Essays on Perception, Performance, Policies, and People

by

Brian Downs

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Management

Dissertation Committee:

Lauren Lanahan, Chair

Chris Liu, Core Member

Natasha Overmeyer, Core Member

Clare Evans, Institutional Representative

University of Oregon

Spring 2025

DISSERTATION ABSTRACT

Brian Downs

Doctor of Philosophy in Management

Title: Essays on Perception, Performance, Policies, and People

This dissertation advances understanding of how supervisors shape equity and performance outcomes in hierarchical organizations. The first study examines partial blindness in promotion decisions, revealing that concealing race and gender has limited effectiveness when supervisors can still infer demographic characteristics through indirect cues. The second study explores how supervisors' behavioral health-seeking behavior affects their appraisal of employees, drawing on attribution theory and human sustainability leadership. The study does not find evidence that supervisors who seek behavioral health support rate employees more generously, but does uncover persistent disadvantages for racial minorities in the sample. Across both studies, the dissertation highlights a critical gap in existing policy design: overlooking the central role of the supervisor. By focusing on the interaction between policy, identity, and managerial behavior, this work contributes a novel lens on meritocracy and accountability in organizational advancement.

ACKNOWLEDGMENTS

I extend my deepest gratitude to the U.S. Army's Medical Service Corps for affording me the privilege to pursue a PhD. I am grateful to the University of Oregon's Lundquist College of Business and its Management Department faculty for welcoming me as a student and inspiring me to strive for excellence.

I am especially thankful to Chris Liu, Natasha Overmeyer, and Clare Evans for their invaluable support and guidance throughout this journey. A very special thank you to my committee chair, Lauren Lanahan, whose unwavering dedication challenged me to meet the highest standards, set clear and actionable goals, and ultimately helped me become a better researcher.

Finally, this achievement would not have been possible without the unconditional love and steadfast support of my amazing wife. She is my rock, my constant source of encouragement, and the foundation that made this journey possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Organizations today face significant challenges in designing policies that balance practical feasibility with the imperatives of fairness and well-being. Some efforts, such as blindness policies, aim to mitigate bias in organizational decision-making by obscuring identity markers (Bertrand and Mullainathan, 2004; Goldin and Rouse, 2000). Others seek to promote human sustainability while maintaining individual and organizational performance (Barnes and Wagner, 2023; Kristof-Brown et al., 2005). Yet these interventions often yield mixed results, especially in hierarchical and culturally entrenched systems. This dual focus—structural fairness and the interpersonal dynamics between leaders and employees—offers a fertile ground for understanding how organizations can build more meritocratic and sustainable workplaces.

This dissertation comprises two complementary investigations. The first study examines whether partial blindness policies in promotion decisions promote meritocracy by obscuring demographic cues such as race and gender. Drawing on tournament theory (Rosenbaum, 1979; Forbes, 1987) and relational demography (Tsui and O'Reilly III, 1989; Tsui et al., 1992), I evaluate which demographic groups, if any, benefit from the partial removal of demographic information during promotion evaluations. My analysis leverages a unique dataset of U.S. Army Of-

ficers, observing promotion outcomes to an upper-level rank between 2017 and 2023 using quarterly snapshots of demographic characteristics, professional roles, supervisory details, and annual performance evaluations. Using a stratified interrupted time series approach, the findings reveal no significant impact of the partial blindness policy on promotion rates for underrepresented groups, lending support to my first hypothesis. For the second hypothesis, I find partial support that varies by demographic group. Women experience increased promotion likelihood when they share higher-than-average racial or gender similarity with their supervisors. For non-white individuals, promotion probabilities improve when they are more frequently evaluated by supervisors within the same occupational workgroup. These findings reveal that partial blindness policies may fail to disrupt entrenched patterns of demographic preference, suggesting that remaining identifiers still enable bias (Younkin and Kuppuswamy, 2018) in promotion decisions. This study contributes to the management literature by demonstrating how partial anonymization policies can fall short of meritocratic ideals, particularly when relational demography continues to influence evaluative outcomes. This raises critical questions about the trade-offs between a policy's feasibility and its effectiveness in translating performance into career advancement (Castilla and Benard, 2010).

The second study turns to the role of supervisors in balancing organizational demands with employee well-being. Specifically, it investigates whether supervisors' own behavioral health (BH) seeking behaviors influence how they evaluate subordinate performance. Drawing on attribution theory (Swift et al., 2013) and human sustainability leadership (Barnes and Wagner, 2023; Luthans and Church, 2002), the study explores whether leaders who engage in wellness-supportive behaviors exhibit greater empathy or reduced bias in appraisal contexts. Using a unique dataset of active-duty U.S. Army officers from 2010 to 2024, I examine supervisor-employee

dyads, distinguishing between the extensive (any BH usage) and intensive (frequency of usage) margins of supervisor behavioral health seeking. In limited support of one of my hypotheses, I find a small effect that when one of the two supervisors' have sought BH, the employees sees a reduced performance rating. However, separate from my hypotheses, I find a persistent performance penalty for racial minorities—despite observable control variables—suggests enduring structural disadvantage (Acosta et al., 2014). These findings reveal the tension supervisors face between organizational performance expectations and human sustainability goals, especially within a system that formally promotes wellness but informally reinforces output-based evaluation (DiBenigno, 2020). This chapter contributes to leadership and performance management literatures by linking supervisors' self-care behaviors to appraisal practices, offering practical implications for fostering equity and sustainability in high-pressure environments.

This research aims to bridge these perspectives by examining how policies and interpersonal relationships can complement one another to advance equity and well-being in hierarchical organizations. By situating blindness policies and behavioral health alignment within the broader context of organizational justice and sustainability, this work provides actionable insights for scholars and practitioners navigating the complexities of modern workplace dynamics.

Practically, the findings underscore the limitations of partial policy interventions and highlight the importance of aligning institutional systems with interpersonal behaviors to achieve lasting change. For organizations seeking to foster fairness and reduce bias, the results suggest that promoting individual wellness among leaders is not sufficient without structural safeguards in appraisal systems. Together, these studies offer a multi-level perspective on building healthier, more just organizations—especially in high-stakes environments like the military, where performance expectations and cultural norms can obscure underlying inequities.

Theoretically, this dissertation advances the literature on performance management and leadership by integrating structural interventions, such as identity-blind evaluations, with relational mechanisms, including supervisor wellness behaviors, to explore how meritocracy is constructed, or constrained, within hierarchical organizations. In the first study, I empirically test the boundaries of partial blindness policies, challenging foundational assumptions in tournament theory that posit merit-based competition as inherently neutral. The findings reveal that structural interventions, while intended to promote fairness, may be insufficient to disrupt entrenched relational biases. This work also extends relational demography by illustrating how employee-supervisor demographic intersections, including race, gender, and occupational alignment, influence promotion trajectories under partial anonymity. These results suggest that even when some demographic cues are obscured during promotion board evaluations, the structural embedding of relational demography in earlier supervisory assessments may perpetuate in-group favoritism and implicit bias, challenging the assumption that partial anonymity ensures meritocratic advancement.

The second study builds on attribution theory by illuminating the role of supervisors' behavioral health engagement in shaping evaluative biases. This contributes to the understanding of leader-driven bias, indicating that personal wellness behaviors may not translate into more equitable appraisals and may, in some cases, reinforce stigmatized perceptions of employee reliability. Moreover, this dissertation bridges the gap between performance management and workplace well-being, two domains that are often studied in isolation. By empirically linking employee performance with relational supervisor behaviors, this research underscores the interconnectedness of evaluative fairness and employee well-being. This integrated perspective contributes to emerging scholarship on human sustainability leadership, illustrating that leader wellness practices

alone are insufficient to address inequities in performance evaluations without concurrent structural changes. Collectively, these findings call for a reexamination of how organizational policies and supervisor behaviors coalesce to shape career trajectories, challenging the assumption that structural interventions alone are adequate to achieve meritocratic advancement.

Chapter 2

FUZZY FAIRNESS: PARTIAL BLINDNESS POLICY EFFECTS ON PROMOTION OUTCOMES

2.1 Introduction

How can organizations ensure that personnel decisions are grounded in fairness and rooted in merit? This question lies at the heart of debates over workplace policies aimed at minimizing bias. Among these policies, "blindness" – where demographic identifiers such as race and gender are concealed – has gained attention for its potential to reduce bias in decision-making processes (Goldin and Rouse, 2000; Krause et al., 2012). While much of the existing literature focuses on hiring outcomes, there is limited research on how blindness policies influence internal promotions. Promotion decisions, which often rely on subjective evaluations and institutional norms, present unique challenges for implementing policies that enhance meritocracy. Humans can be biased in organizational decision-making even with minimal group membership information (Bertrand and Mullainathan, 2004; Goldin and Rouse, 2000; Younkin and Kuppuswamy, 2018), and tend to compensate for missing information by overemphasizing other available details (Goldin and Rouse, 2000; Norton et al., 2004; Younkin and Kuppuswamy, 2018). By re-

moving visible group membership cues, organizations aim to level the playing field, ensuring that hiring, promotion, and other decisions are based solely on qualifications, performance, and merit.

Promotion processes in up-or-out organizations, such as the military or law firms, can be conceptualized as sequential tournaments where members of a cohort compete for limited advancement opportunities. Classic tournament mobility theory posits that early career advancements give lasting advantages in subsequent competition rounds.[1] By structuring promotions as relative performance contests, organizations seek to incentivize high performance. Winners earn substantial rewards, like a promotion, increased pay, or higher status, motivating all employees to exert effort even when only a few will advance each round (Rosenbaum, 1979). In theory, a meritocratic tournament yields efficient outcomes where the most capable individuals rise to the top. However, this idealized outcome assumes that promotions are decided purely on performance. While in practice, other factors such as demographic similarities between evaluators and candidates can significantly influence advancement decisions (Castilla and Benard, 2010; Rivera, 2012).

In hiring actions, many examples show how women and racial minorities see increased callback rates in hiring when blindness or anonymous job application procedures are used (Bertrand and Mullainathan, 2004; Krause et al., 2012; Yavorsky, 2019). However, administering these policies for extended periods of time is often unsustainable or even unfeasible to implement in the first place (Public Service Commission of Canada, 2018; Eliezer et al., 2024). There could be burdensome monetary costs needed for upfront investments in software, storage, and compliance

[1]Tournament theory is an economic concept developed by Lazear and Rosen (1981) which explains how compensation structures based on relative ranking rather than absolute productivity can motivate performance. The fundamental principle is that workers compete for predetermined prizes (such as promotion or increased compensation) in a rank-order competition, where the attribution of these prizes depends on the position an employee's production occupies in comparison to others, not on its absolute value. This approach is particularly efficient when quantifying output is difficult or expensive, but ranking workers is relatively easy (Connelly et al., 2014).

monitoring equipment to convert and store records in new information formats. Implementing new processes requires overcoming organizational resistance and institutional inertias, delaying decision-making processes. Finally, developing new processes and institutional training courses would be essential to standardize the understanding and implementation of policies across the organization.

Given the evidence of performance-reward bias in managerial decision-making for performance appraisals and salary growth (Castilla, 2008), I explore whether partial blindness policies can mitigate these biases and promote fairness in organizational advancement. Specifically, how do partial blindness policies affect performance-based advancement in up-or-out promotion systems, advancing from middle to senior management, where full blindness policies are impractical? Full blindness policies, while effective in initial hiring decisions, are often unfeasible for internal promotions due to the need for supervisors to assess leadership potential, job-specific expertise, and past performance. These are evaluations that inherently require some knowledge of the employee's identity and work history. In contrast, partial blindness policies attempt to strike a balance between anonymity and practicality by masking select demographic information (e.g., race, gender, and marital status), while retaining identifiers such as names and pronouns.

While much of the existing literature focuses on hiring outcomes, there is limited research on how blindness policies influence internal promotions, which present unique challenges. Unlike entry-level hiring, promotion decisions often rely on subjective evaluations from supervisors who have ongoing relationships with candidates and access to detailed performance histories. As Forbes (1987) demonstrated, later career events and specific job assignments can significantly influence ultimate career attainment. In up-or-out promotion systems, this means employees must not only compete with peers, but the quality of opportunities and mentorship they receive early

on can set them on different trajectories, potentially limiting the effectiveness of later intervention through blindness policies.

By analyzing how partial blindness policies impact performance-based advancement within organizational promotions, my study offers a nuanced understanding of how masking *some* demographic information (e.g., race, gender, and marital status), but not other information (e.g. names and pronouns), affects promotion rates for individuals in organizations with up-or-out promotion systems. My findings contribute to current research by providing insights into how organizations can strike a balance between policy effectiveness and practicality in fostering diversity.

The primary data for this study is individual level personnel files and annual performance appraisals of a large governmental organization from 2017 to 2023. My primary model is a stratified interrupted time series (SITS) estimation between pre- and post- policy cohorts of employees that are trying to advance from the rank of middle to senior management. The dependent variable is the promotion of the individual to the senior management level. The primary independent variable is the implementation of the partial blindness policy in 2020. To ensure my findings are not sensitive to my model specifications, I conduct a robustness check using a linear probability model. I find evidence that this type of partial blindness policy, where group membership information is concealed, but names and pronouns are not hidden, produced inconclusive results regarding its ability to enhance meritocratic decision-making for promotions.

Despite the well-intentioned design, partial blindness policies, which obscure some demographic markers like race and gender while leaving others (such as names and pronouns) visible, may inadvertently perpetuate the very biases they aim to eliminate. By occupying a "middle ground," these policies still allow evaluators to infer key demographic characteristics, subtly rein-

troducing bias through residual identifiers. This incomplete anonymization may enable decision-makers to engage in casuistry. That is the selective interpretation of qualifications to justify pre-conceived evaluations, under the guise of impartiality (Norton et al., 2004). From a tournament theory perspective, partial blindness policies fall short of correcting for cumulative career advantages that are often entrenched early on, preserving hierarchical structures that benefit dominant groups. In this way, policies designed to neutralize bias may paradoxically entrench it, complicating efforts to achieve genuine meritocracy in up-or-out promotion systems.

This study incorporates several important extensions since my dissertation proposal. First, I explored how early entry and placement into supervisory roles influence long-term promotional outcomes by tracking the number of employees an individual has supervised, the duration of their supervisory roles, and the timing of their initial supervisory placement. Next, I introduced a second hypothesis that examines demographic similarity in the employee-supervisor dyad based on race, gender, and occupational grouping. Finally, to demonstrate additional robustness, I reran the analysis using Coarsened Exact Matching. Collectively, these extensions build upon and advance prior research on career trajectories within tournament-style promotion systems (Forbes, 1987; Rosenbaum, 1979), offering new insights into how demographic alignment and early supervisory experience shape promotion outcomes.

This study contributes to the management literature by examining the effects of partial blindness policies on merit-based advancement in a large firm. By exploring whether promotion decisions are shaped by entrenched informal advantages, such as networks or early mover status, rather than formal evaluation metrics, this study highlights the potential limitations of a single policy in fostering greater meritocracy. The empirical findings provide a nuanced understanding of these dynamics and offer valuable insights for human resource professionals, executives, and

policymakers seeking to design more equitable promotion systems.

2.2 Literature Review

The literature review is organized into four main sections. First, I explore the persistence of bias and perception in the labor market, highlighting how demographic factors such as race and gender influence workplace outcomes. Next, I examine the effectiveness of blindness policies in mitigating these biases, with a focus on their application in hiring and promotion contexts. Then, I delve into relational demography and its role in shaping promotion outcomes, emphasizing how demographic similarities between supervisors and subordinates can influence evaluations. Finally, I discuss the theoretical underpinnings of partial blindness policies and their potential limitations, setting the stage for the hypotheses development and empirical analysis that follow.

2.2.1 Bias and perception in the labor market

It is naïve to say that people can be blind to race, considering how early and quickly the distinction is made in individuals (Apfelbaum et al., 2012; Bar-Haim et al., 2006; Ito and Urland, 2003). Unfortunately, observing race or a person's skin color is innate to our human condition. It occurs as early as six months of age, and the length of time needed for that perceptual differentiation of race to occur is in less than one-seventh of a second (Apfelbaum et al., 2012; Bar-Haim et al., 2006; Ito and Urland, 2003). Extant literature already shows that racial and gender bias persist in the labor market. Studies have consistently shown that Black, Hispanic, and female workers face wage and promotion disparities (Cancio et al., 1996; Greenman and Xie, 2008). Occupational segregation reduces opportunities for promotion and limits access to workplace benefits, especially for women of color (Cotter et al., 2003; Snipp and Cheung, 2016). These biases infiltrate

the labor market from the outset, as shown in job hiring research.

Studies suggest that factors such as gender, age, and ethnic background influence the likelihood of receiving an interview invitation (List, 2001). Research shows that job applicants with black names are less likely to receive interview callbacks compared to those with white-sounding names (Bertrand and Mullainathan, 2004), similar to patterns observed in Germany and the Netherlands (Krause et al., 2012; Quillian et al., 2017). Gaddis (2015) further highlights how race and college selectivity intersect in the labor market, revealing that black applicants with racially identifiable names receive fewer callbacks, even when holding degrees from elite universities, compared to white applicants.

Younkin and Kuppuswamy (2018) further reveal how racial bias extends beyond hiring, showing that black founders in crowdfunding platforms face significant disadvantages compared to white founders, even when presenting projects of identical quality and performance, highlighting the pervasive nature of race-based discrimination across different decision-making contexts. As a final point, research shows when racial and gender information is included in hiring and performance actions, that information can penalize equally qualified, capable, or performing women and minorities compared to the majority group (Åslund and Skans, 2012; Benson et al., 2023; Castilla, 2008; Krause et al., 2012; Landau, 1995). Given these disparities, organizations aiming to increase diversity may need to adopt a different approach to reduce the visibility of group membership.

2.2.2 Blindness and its effectiveness in labor markets

In response to these biases, organizations have adopted blindness policies aimed at reducing the influence of demographic markers and group membership information. Blindness policies are

designed to ensure that group membership (e.g., race, gender, sexual orientation) does not influence decision-making, perceptions, or actions (Apfelbaum et al., 2012). Implemented across educational, organizational, legal, and political domains, these policies aim to downplay gender, racial, and ethnic differences, fostering unity through common goals and values (Apfelbaum et al., 2010; Butler and Denton, 2021).

For instance, orchestra auditions behind screens led to increased female selection (Goldin and Rouse, 2000), while anonymous applications in France and Sweden improved outcomes for women, though racial bias remained (Åslund and Skans, 2012; Behaghel et al., 2015). A 2004 resume experiment found that white names received 50 percent more callbacks for interviews than similarly qualified black applicants (Bertrand and Mullainathan, 2004), with a similar finding in racial call back rates in a Dutch City in 2011 (Bøg and Kranendonk, 2011). Generally, speaking these results suggest that racial and ethnic discrimination may be harder to circumvent than gender discrimination (Åslund and Skans, 2012).

The literature on blindness policies highlights significant challenges of partial implementations, particularly in how bias can infiltrate processes, such as name-based bias. Research suggests that a color-blind approach may obscure racial discrimination, reducing the likelihood of detecting and reporting overt injustices (Apfelbaum et al., 2010). That is the effects of discrimination persist, even though the policy is intended to reduce it.

While many studies have focused on the effectiveness of full blindness policies in improving diversity during the hiring process, much less is known about how these policies influence internal promotions where full blindness is less feasible. Implementing fully blind promotion systems in large bureaucracies faces substantial institutional hurdles. Eliezer et al. (2024) noted in her study with the Department of Defense (DOD) that the feasibility of removing all identifying

information (names, pronouns, etc.) from promotion board files was very low given current systems and procedures. In one pilot program, the labor-intensive nature of redacting thousands of existing personnel records presents significant compliance risks and resource costs (Public Service Commission of Canada, 2018), let alone the millions of records required by the DOD. Given these constraints, partial blindness is seen as more attainable—removing photographs and obvious demographic markers while retaining names and pronouns in evaluation narratives (often containing gendered language). However, this compromise approach may allow bias to persist through residual cues and previously biased evaluations.

2.2.3 Relational demography and bias in promotion

Compared to hiring, promotion processes are inherently more complex, shaped by organizational norms, informal networks, and individual variables (Markham et al., 1987). While organizations may outline official promotion pathways, advancement opportunities are often constrained by vacancy rates, internal labor market structures, and seniority-based career ladders. Also, the organization's social structure, shaped by informal networks, sponsorship and managerial discretion can help employees develop strong relationships or gain visibility in key roles that lead to greater mobility compared to employees who have weaker networks. Lastly, subjective assessments such as leadership potential and prior role assignment can play a significant role in determining advancement (Markham et al., 1987).

Relational demography theory posits that demographic similarities between supervisors and subordinates significantly influence workplace evaluations. When supervisors and their subordinates share demographic attributes, increased interpersonal attraction and communication ease can unconsciously bias evaluations in favor of similar subordinates (Tsui and O'Reilly III,

1989; Rivera, 2012). This dynamic is particularly relevant to the hypotheses in this study, as it suggests that even under partial blindness policies, residual demographic cues, such as names or pronouns, may allow decision-makers to infer group membership and continue exhibiting in-group preferences.[2] Consequently, these biases could persist in performance evaluations and promotion decisions, undermining the intended meritocratic outcomes of partial blindness policies.

In promotion systems based on performance appraisals, these similarities can create meaningful advantages—officers sharing race or gender with evaluators may receive stronger ratings and greater sponsorship than equally qualified colleagues from different demographic groups. These advantages compound over time, determining who accesses key developmental opportunities and ultimately advances in rank. This homophily-based bias undermines meritocratic promotion tournaments and may persist despite partial blindness policies, as biased evaluations from earlier career stages remain embedded in candidates' records when they reach promotion boards.

Additionally, the composition of promotion boards, clarity of performance criteria, and existing diversity initiatives can all condition policy success. If a board's members are homogeneous, even under partial blindness they may share similar preferences in what constitutes a strong record (Thomas and Wise, 1999). In contexts where performance criteria are subjective—such as leadership potential—bias has more room to operate (Tversky and Kahneman, 1974; Castilla and Benard, 2010). Some research indicates that demographic disparities in promotions often trace back to earlier pipeline stages rather than bias in board deliberations alone (Eliezer et al., 2024; Hesli et al., 2012), suggesting that interventions at the final selection stage may have diminished impact if upstream inequities remain unaddressed.

[2]The Department of the Army does keep a record of the race and gender composition of its promotion board members and how they score each employee considered for promotion. However, I am unable to access this data, and it is my understanding from talking to anonymous sources that promotion board voting data is some of the most closely guarded data within the Office of the Secretary of the Army.

2.2.4 Hypotheses Development

I argue that partial blindness policies, where key demographic indicators like names and pronouns remain visible, but other demographic details like race and gender remain hidden, are unlikely to yield significant improvements in diversity outcomes. This "middle ground" still leaves space for bias to creep in, allowing individuals to infer race and gender through remaining identifiers, and engage in casuistry—manipulating qualifications to justify biased decisions (Norton et al., 2004). Partial blindness may inadvertently reinforce biases by presenting decision-makers with just enough information to influence their judgments, while still maintaining an illusion of impartiality.

Research consistently shows that diversity contributes to innovation, creativity, and problem-solving (Ahuja and Katila, 2001; van Dijk et al., 2012). Without blindness policies or effective alternatives, organizations may miss the social and economic benefits of a more diverse workforce. For example, diverse teams bring varied perspectives, enabling higher levels of innovation and creativity, which are essential in competitive and dynamic markets (van Dijk et al., 2012). Thus, discriminatory practices undermine these benefits, keeping the workforce less diverse and limiting organizational performance.

In theory, removing demographic cues or group membership information should improve meritocratic decision-making. However, evidence suggests that meritocracy often fails to prevail in practice, particularly in promotions. Organizations with stated meritocratic ideals have been shown to favor male employees over equally qualified female counterparts (Castilla, 2008), and gender-blind models in science reveal unequal career trajectories between men and women due to societal and familial pressures (González Ramos et al., 2015; Graddy-Reed et al., 2019). These

findings imply that partial blindness policies, which obscure only some demographic information, are unlikely to reduce these entrenched biases.

Building on tournament theory and relational demography research, I argue that partial blindness policies are unlikely to meaningfully improve equity in up-or-out promotion systems, particularly when demographic cues such as names and pronouns remain visible. In such systems, promotion decisions are shaped by cumulative evaluations over time, and early career advantages can lock in trajectories before promotion boards even convene (Rosenbaum, 1979; Forbes, 1987). Because performance appraisals—key inputs into board decisions—are completed prior to the implementation of any blindness, biases rooted in supervisor evaluations remain embedded in the record (Tsui and O'Reilly III, 1989). Therefore, any attempt to enhance meritocracy at the board level through partial anonymity may come too late to offset disparities already baked into candidates records.

Hypothesis 1: In organizations with up-or-out promotion systems, a partial blindness policy will not alter promotion outcomes based on individual performance metrics, compared to a promotion system with no blindness policy. (H1)

The effectiveness of blindness policies may be moderated by several organizational factors. Early career role assignments represent a significant potential moderator, functioning as initial "seeding" of competitors. Forbes (1987) observed that initial functional areas and job variety were significant predictors of ultimate career success, operating as signals to promotion decision-makers about candidates' capabilities. In hierarchical organizations like the military, differential early career opportunities can moderate blindness policy effects—when promotion files reflect

accumulated advantages from prior role allocations that were themselves influenced by biases or favoritism, concealing identity at the selection stage may do little to change comparative outcomes (Rosenbaum, 1979; Lytell et al., 2023).

More specifically, relational demography suggests that demographic similarity between supervisors and subordinates systematically influences performance ratings and promotion outcomes (Tsui et al., 1992; Castilla, 2008). When remaining identifiers, such as names or pronouns, permit inference of group membership, decision-makers may continue to exhibit in-group preferences even under partial blindness. If these biases influence earlier evaluations, they can persist into final promotion decisions regardless of later anonymization efforts.

Hypothesis 2: Under a partial blindness policy, promotion outcomes will continue to favor candidates who share demographic similarities with decision-makers, as inferred through remaining visible identifiers like gender, race, or occupational workgroup. (H2)

2.3 Methods

2.3.1 Empirical Context

The sample for this study is a panel of United States Army (hereafter, Army) officers serving between 2017 and 2023. The Army provides a unique context for this study due to its up-or-out promotion system, where serving at lower levels of the organization is required to advance to higher levels of the organization. This offers a controlled field to observe promotions over time without lateral entries or external hires that could contaminate the sample. Other organizations that implement similar up-or-out structures include airline, legal, accounting, consulting, govern-

ment, and other military organizations.

Officers' eligibility and selection for promotion occurs based on the amount of time they have served at their current rank, their performance compared to their peers, and the number of positions that need to be filled at the next rank. The primary model of this study focuses on officers' promotion from the rank of major to lieutenant colonel.[3] This transition is important for a few reasons. Lieutenant colonel is the rank where leaders can serve in the pinnacle organizational leadership positions in the Army. It is also a transition rank to more organizationally strategic and executive responsibilities. Additionally, promotion to lieutenant colonel is highly competitive, with fewer available positions compared to lower ranks. This competition might amplify existing organizational biases in performance ranking and decision-making (Castilla, 2011), making promotion to lieutenant colonel an ideal choice to study. While there may be fewer positions compared to lower ranks, there are still enough officers in large enough quantities, across the Army, to allow for robust analysis and statistical conclusion validity. Lastly, lieutenant colonel is the terminal career rank that most officers aspire to achieve and retire at (DoD Office of the Actuary, 2023).

An officer usually serves four to six years at the rank of major before being considered for promotion to lieutenant colonel. Each year, officers receive an annual performance appraisal, where the officer's first-line supervisor, the "rater," evaluates the officer's performance in their current job. A second-line supervisor, the "senior rater," evaluates the officer on their potential to serve at higher ranks and positions of greater responsibility within the organization. The officer being evaluated receives a quantitative ranking from each of their supervisors on a four-tier scale

[3] An officer that has achieved the rank of major has already been promoted three times within the Army (starting as a second lieutenant $\rightarrow$ 1st promotion: to first lieutenant $\rightarrow$ 2nd promotion: to captain $\rightarrow$ 3rd promotion: to major). Promotion to lieutenant colonel marks the fourth career promotion as an officer. The average officer promotes to lieutenant colonel at 17 years of active federal military service as an officer.

and a qualitative narrative about their performance and potential, respectively.[4]

When an officer is considered for promotion, their annual performance evaluations, a one-page curriculum vitae, and other administrative documents are submitted to a promotion board. Figure 1 shows the steps of the promotion process. The requirements of the minimum composition of officers serving on a promotion board are specified in law,[5] and requires that there are at least five officers that serve on the board (Military Leadership Diversity Commission, 2010). This promotion board evaluates each officer up for promotion through a special ranking process (Taliaferro, 2015). Once that process is complete, the promotion board provides an ordered merit list, with a slate of officers recommended for promotion, through the Department of Defense to the US Senate, for approval (US Army, 2020). Once the promotion list is confirmed, the officers are promoted in sequence based on the promotion list. The process from when the promotion board meets to the individual being promoted usually takes between 12 and 18 months to complete.

On July 15th, 2020, then Secretary of Defense Mark T. Esper,[6] issued a memorandum that said the Department of Defense was going to start three initiatives to tackle discrimination, prejudice, and bias in all ranks of the Armed Forces (Esper, 2020). The memo stated that each of the three efforts had the goal of identifying, "actions the Department can take within policies, programs and processes to improve diversity, inclusion and equal opportunity for all Service members" (Esper, 2020).

Starting on August 1st, 2020 the Army began removing photographs, race, gender and

[4]More information on the Army evaluation process can be found in its governing document, Army Regulation 623-3, *Evaluation Reporting System*, published in 2019.

[5]The governing laws for the United States Armed Forces are in Title 10 of the United States Code. The reference specified is United States Code, Title 10, Section 612: Composition of selection boards (*Composition of Selection Boards: 10 U.S.C, Continuous*)

[6]The Secretary of Defense is a civilian (non-uniformed) who serves as the senior official and policy maker in the Department of Defense and is the second in command authority over the U.S. military to only the President.

ethnicity data from promotion board files (Department of the Army, 2020). The Secretary and Chief of Staff of the Army at the time said the purpose was to, "ensure that selection boards are as fair and impartial as possible" (Minitrez, 2020). However, this policy did not remove names or pronouns from narrative statements in the officer's performance evaluation documents, which are shown to promotion boards because it would have been prohibitively expensive from an administrative standpoint with little expected gain (Eliezer et al., 2024).

While on the surface, this appears to tackle issues of bias in promotion selection, it only does so at the end of the process, not the beginning (Prescott, 2020). The annual performance appraisals, which promotion boards heavily rely on, are not conducted blindly. First- and second-line supervisors are aware of the rated officer's personal and professional demographic details—such as marital status, race, gender, and appearance—that are hidden from the promotion board.[7] Although ratings are meant to be objective, each officer can have unique specified and implied responsibilities, making their final evaluation both a reflection of their individual performance and a comparison to their peers. Additionally, the highest performance rating is a zero-sum contest for first- and second-line supervisors. To enforce differentiation in rankings, the Army requires supervisors to limit the highest ranking to fewer than 50% of officers they evaluate at each rank throughout their careers.

Focusing on the act of promotion and the implementation of the partial blindness policy, this study employs a stratified interrupted time series (SITS) design to evaluate the effects of partial blindness policies on promotion outcomes within a military context. This design is advantageous because it approximates a natural experiment. The decision to implement the partial blind-

[7] As a point of clarification, this study uses demographic data of first- and second-line supervisors based on deidentified matching through performance appraisals. Although Army promotion boards typically include diverse representation by gender and race (Eliezer et al., 2024), the specific demographic composition of the promotion boards is not available for this study.

ness policy was directed by the civilian personnel who control the military. This reduces selection bias because the officers being observed had no control over when the policy was implemented or that it was effective across the entire organization simultaneously (Esper, 2020; McInnis, 2020; Minitrez, 2020). Additionally, inclusion into the pre- and post-policy groups is based on when the officers entered military service and were promoted to major, which occurred years before deliberation or implementation of the policy. By comparing within-group promotion rates before and after the policy change, I can observe how the partial blindness policy intervention influences outcomes across different stratified demographic groups. Additionally, the sample is stratified by race (white or non-white), gender (male or female), and promotion cohort. [8]

To improve model specification, I control for individual characteristics related to personality and performance. These include unobservable factors that the promotion board is blind to, such as race, gender, marital status, and age at promotion to major. As well as observable factors like supervisory experience, years of service, military education levels, annual performance appraisal rankings, occupational groups (Army basic branch), and job classification codes (military occupational specialty). These measures provide crucial insights into how partial blindness policies affect promotion dynamics in a highly structured and hierarchical organization like the military. Understanding how demographic factors relate with performance evaluations and institutional needs can reveal whether these policies truly mitigate bias or allow deeper systemic inequalities to persist. By examining these stratifications, I evaluate whether partial blindness policies can effectively create meritocratic opportunities or if additional measures are needed to achieve equitable outcomes regardless of individual demographic characteristics.

[8]The employee grouping for this study includes four promotion cohorts: core operations employees; intelligence and communications employees; administrative, human resources, and logistics employees; and, medical administration employees and nurses.

2.3.2 Data

Data for this study comes from three proprietary DOD databases hosted by the Person-Event Data Environment (PDE). First, The Defense Manpower Data Center (DMDC) personnel master file (hereafter, "master file") provides personal and professional information with quarterly snapshots, at the individual level, for the duration of an officer's military career. This includes a unique individual identifier, rank, race, gender, age, marital status, years of active federal military service, and other individual personal and professional information. The second and third databases, respectively, contain performance evaluation data from the Army's Evaluation Entry System and Officer Evaluation Report database. These databases contain the unique individual identifier of the rated officer used for matching to the master file, the rated officers branch and occupational specialty, the unique identifiers, and quantitative rankings of the first- and second-line supervisors, and the dates of the evaluation period.

2.3.3 Dependent Variables

The dependent variable, $Promotion_i$, is a binary outcome of whether an individual in the primary sample was promoted from the rank of major to the rank of lieutenant colonel (1 = promoted, 0 = not promoted). The data for this variable comes from the master file and is recorded in the first observation in the master file where the individual's rank has changed from major to lieutenant colonel. Access to promotion decision data represents a significant methodological advantage in studying career advancement. Unlike more commonly used proxies in the management literature this dataset offers a direct measure of organizational selection for promotion. Selection decisions are central to organizational mobility (Rosenbaum, 1979) and power dynam-

ics (Sheridan et al., 1990), yet they remain largely unexplored in empirical research due to data accessibility constraints. This unique access not only enhances the robustness of the findings but also contributes to bridging a critical gap in the empirical examination of organizational advancement. Furthermore, because the data are drawn from a structured organizational promotion processes, the analysis benefits from consistency in evaluation criteria and centralized decision-making, reducing unobserved variability often present in private-sector datasets.

2.3.4 Key Explanatory Variable

My key explanatory variable is exposure to the partial blindness policy, $Post_t$. While the policy was officially implemented in 2020, the effects on promotions would not be observable until 2021. Therefore, for comparison, I define the pre-policy period (coded as post = 0) as the three-year promotion window from 2018-2020, and the post-policy period (coded as post = 1) as the three-year promotion window from 2021-2023. Because all individuals eligible for promotion during the defined windows are included, placement in the pre- or post-policy sample is determined solely by promotion timing, leaving no opportunity for manipulation between groups.

2.3.5 Controls

In addition to the primary variables, this study includes several control variables to account for confounding effects. The full list of variables, their operationalization, and their data source can be found in Table A.1. At the individual level, key personal characteristics such as gender, race, and age are considered. For individual professional characteristics, variables include level of service and joint professional military education and previous supervisory experience at any point in their career.

I control performance by creating two variables. The first variable is an individual's average second-line supervisor score for their annual performance appraisals while at the rank of major. This average score is created by summing the evaluation ranking (scale: 1-4, higher is better) given for each evaluation by the individual's second-line supervisor and then dividing by the total number of evaluations received while serving at the rank of major. The second variable is the composite average of the first- and second-line supervisor's evaluations for an individual, created by summing the evaluation ranking given for each evaluation by both supervisors and then dividing by the total number of evaluations received while serving at the rank of major (scale: 2-8, higher is better). Both performance scores are then standardized for inclusion in the regression to allow for ease of comparison and interpretation. Lastly, I control for professional opportunities by accounting for Army functional work groups and job classification codes, recognizing that career tracks vary in advancement opportunities.

To account for relational demography, I include binary indicators for whether the employee was of the same race, gender, or occupational workgroup as their first-line supervisor and second-line supervisor, respectively. This is done to account for the potential bias that may occur in the performance evaluation process for supervisors that show preference for employees that are more similar to themselves.

Lastly, I created three variables to serve as indicators of career trajectory compared to peers. First, this includes how many years before being eligible for lieutenant colonel the individual first became a supervisor of other officers. Second, this includes how many years the individual has been a supervisor of other officers. Lastly, this includes how many individuals the officer has supervised in their career, up to being eligible for promotion to lieutenant colonel. These variables are included to account for the potential bias that may occur in the performance evalu-

ation process for supervisors that show preference for employees that have had more supervisory experience or have supervised more individuals than their peers.

2.3.6 Data Structure

The primary model of this study examines two cohorts of Army majors at the individual level. In my sample, individuals are assigned exclusively to the pre- or post-policy groups based on the year they were promoted to major: the pre-policy group achieved the rank of major between 2011 and 2014, with promotion to lieutenant colonel between 2018 and 2020, while the post-policy group reached major between 2015 and 2018 and advanced to lieutenant colonel between 2021 and 2023. Following the approach used in other studies of the military, I control for individual aspiration by treating all individuals eligible to be promoted as desiring promotion, thereby reducing bias from motivational differences (Eliezer et al., 2024; Lytell et al., 2023).

The master file was precisely matched with performance data. The officers included in the analysis were active-duty Army officers from occupational groups entered directly by entry-level officers, and excluding U.S. Army National Guard and U.S. Army Reserve personnel. Additional observations were excluded if their promotion data was incomplete, contained extraneous values, did not have a job classification code listed while at the rank of Major, had no observations for years of service, or had a null value for observed marital status. Observations were removed if individuals exited the data before the start of the promotion window to lieutenant colonel, addressing potential censoring issues and ensuring a complete analysis of individuals eligible for promotion (Lytell et al., 2023).

Table 2.1: Descriptive statistics for partial blindness policy promotions

Variables	Count	Mean	SD	Min	Max	Mean (Pre-Policy)	Mean (Post-Policy)	t-statistic	p-value
Promoted to LTC	10595	0.861	0.346	0.000	1.000	0.879	0.840	5.812	0.000
Males promoted to LTC	8808	0.858	0.349	0.000	1.000	0.879	0.834	6.058	0.000
Females promoted to LTC	1787	0.872	0.334	0.000	1.000	0.878	0.866	0.777	0.437
Whites promoted to LTC	7390	0.876	0.330	0.000	1.000	0.896	0.853	5.607	0.000
Non-whites promoted to LTC	3205	0.826	0.379	0.000	1.000	0.840	0.811	2.146	0.032
All individuals, exlcuding white males	10595	0.389	0.488	0.000	1.000	0.377	0.403	-2.731	0.006
All not white individuals	10595	0.303	0.459	0.000	1.000	0.299	0.307	-0.902	0.367
All females	10595	0.169	0.374	0.000	1.000	0.157	0.182	-3.505	0.000
Age when promoted to Major	10595	37.581	4.001	29.000	57.000	37.508	37.662	-1.979	0.048
Individuals married	10595	0.883	0.322	0.000	1.000	0.883	0.882	0.125	0.901
Employee and 1LS are same gender (binary)	10595	0.503	0.500	0.000	1.000	0.522	0.482	4.166	0.000
Employee and 1LS are same race (binary)	10595	0.405	0.491	0.000	1.000	0.423	0.386	3.844	0.000
Employee and 1LS are same occupational workgroup	10595	0.295	0.456	0.000	1.000	0.291	0.300	-1.016	0.310
Employee and 2LS are same gender (binary)	10595	0.526	0.499	0.000	1.000	0.554	0.496	5.965	0.000
Employee and 2LS are same race (binary)	10595	0.428	0.495	0.000	1.000	0.448	0.406	4.310	0.000
Employee and 2LS are same occupational workgroup	10595	0.231	0.422	0.000	1.000	0.241	0.220	2.484	0.013
Years of AFMS when eligible for promotion to LTC	10595	19.605	3.264	6.000	37.000	19.512	19.708	-3.096	0.002
Officer has completed ILE or SSC	10595	0.958	0.202	0.000	1.000	0.953	0.962	-2.213	0.027
Any Joint Professional Military Education	10595	0.108	0.310	0.000	1.000	0.111	0.104	1.154	0.249
Years from first rating officers to eligibility for LTC	10595	0.754	2.304	0.000	17.000	0.578	0.949	-8.303	0.000
Years as a supervisor of officers	10595	0.695	2.105	0.000	12.000	0.667	0.725	-1.414	0.157
Number of officer evaluations as a supervisor	10595	10.467	35.326	0.000	568.000	9.911	11.081	-1.703	0.089
Average Senior Rater Score (Scale:1-4)	10595	3.619	0.280	2.192	4.000	3.603	3.637	-6.261	0.000
Average Score of Rater and SR (Scale: 2-8)	10595	7.380	0.459	4.538	8.000	7.414	7.341	8.244	0.000
Observations by Policy Group	10595					5558	5037		

Note: The table presents descriptive statistics for the full sample (N=10,595). The group means are identified by if the employees primary promotion window occurred before or after the partial blindness policy went into effect. AFMS = Active Federal Military Service. SD = Standard Deviation. 1LS = First Line Supervisor (Rater). 2LS = Second Line Supervisor (Senior Rater).

2.3.7 Descriptive Statistics

Table 2.1 presents descriptive statistics for the 10,595 observations in the sample. The pre-policy period consists of 5,558 observations and the post-policy period includes 5,037 observations. The first set of columns reports for the full sample of officers, including those in the pre- and post- period (col. 1-5). The second set of columns (col. 6-9) report statistics for the comparison of means in all measures between the pre- and post-policy groups.

In the full sample, on average, promotion to lieutenant colonel is achieved by approximately 86 percent of the sample. Approximately 61 percent of the sample is white males, approximately 30 percent of the sample is not white, and approximately 17 percent of the sample is female. Overall, all demographic groups see a decrease in promotion rates between the pre- and post-periods. This can be attributed to a combination of Army requirements for promotion based on organizational structure, reduced attrition at higher ranks, an increased officer pool at

key ranks, or budgetary constraints constraining the number of personnel that should advance. There is a smaller rate of promotion decrease among females; however, females achieve a 2.5% increase in representation in the post-period.

Across the other explanatory variables, the pre- and post- policy groups are generally comparable. For the standardized composite performance score, the pre-policy period has a higher value because the data contains a four-year period from 2011 to April 2014, inclusive, where raters did not have to restrict the number of their top rankings of officers to less than 50 percent, skewing the mean for those years. However, this does not impact the outcomes because this performance policy was applied uniformly to all employees and their supervisors in their respective promotion year-group cohorts (Dalen, 2014).

2.3.8 Research Design

To analyze the impact of the partial blindness policy for different demographic group promotions, I used a SITS model to examine pre- and post-policy promotion data for Army officers advancing from major to lieutenant colonel. This quasi-experimental approach stratifies the sample, represented by group W, by gender (male and female), race (white and non-white), and promotion cohort and the model is run separately with each group. For inclusion criteria, I determine each officer's year of promotion to major. Equation 2.1 presents the SITS logistic regression model in cross-sectional form, with i representing individual officer observations.

2.3.9 Model Specification

$$\begin{aligned}\text{logistic}(\text{P}(\text{Promotion}_\text{i} = 1|\text{Group} = \text{W})) = \beta_0 + + \beta_1(Post_t) \\ + \beta_2(Not.white_i) + \beta_3(Female_i) \\ + \beta_4(Relational.Demography_{ij}) \\ + \beta_5(Supervisory.Experience_{it}) \\ + \beta_6(Military.Education_i) \\ + \beta_7(Performance_i) + \beta_8(Controls) + e_i\end{aligned} \tag{2.1}$$

2.4 Results

Table 2.2 and Figure 2.1 report the results of equation 2.1, estimating the intervention effect of the promotion policy on promotions from major to lieutenant colonel for the complete set of individual demographic groups. The impact of the policy on promotion rates across the full sample of 10,595 observations results in an approximately 5 percent increase in probability of promotions ($\beta = 0.051, SE = 0.011$). The results are comparable when evaluating the policy for white ($\beta = 0.049, SE = 0.011$) and non-white officers ($\beta = 0.047, SE = 0.022$), collectively. When the results are broken out by race and gender (Table 2.3) the effect remains for white men ($\beta = 0.051, SE = 0.012$), non-white men ($\beta = 0.065, SE = 0.025$), and white women ($\beta = 0.046, SE = 0.023$). For women broadly, and non-white women specifically, the results are inconclusive. Therefore, within this instance, the implementation of a partial blindness policy did not uniformly alter promotion outcomes across demographic groups.

Further analyzing the results, I see significant effects in places I would expect in the full

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due to the Partial Blindness Policy

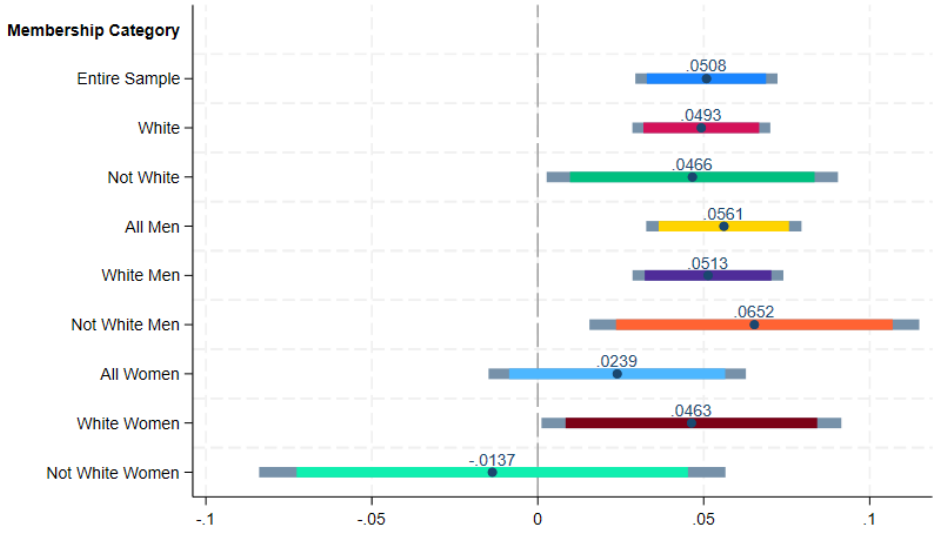

Figure 2.1: Primary Results of the SITS Model

sample for employees to advance in this organization. This includes significant effects for completing required professional schooling (Intermediate Level Education (ILE) or Senior Service College (SSC)) ($\beta = 0.126, SE = 0.018$), getting higher scores from your second-line supervisor ($\beta = 0.184, SE = 0.022$), and overall performance score from your first and second-line supervisors ($\beta = 0.181, SE = 0.013$).

In table 2.3, an interesting finding is how white and non-white women have effects in different directions based on the total number of evaluations (logged) they have completed as a supervisor. White women see a penalty for probability of promotion as the number of evaluations they conduct increases ($\beta = -0.128, SE = 0.047$). While non-white women see an increase in promotion potential as the number of evaluations increase ($\beta = 0.122, SE = 0.058$).

Table 2.2: Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, by race or gender

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo All Whites	(3) Logit ITS dv_promo All Non-Whites	(4) Logit ITS dv_promo All Men	(5) Logit ITS dv_promo All Women
Post Policy: post == 1 = 1	0.051***	0.049***	0.047**	0.056***	0.024
	(0.011)	(0.011)	(0.022)	(0.012)	(0.020)
All not white individuals = 1	-0.008			-0.005	-0.006
	(0.010)			(0.008)	(0.027)
All females = 1	0.012	0.016	0.012		
	(0.014)	(0.018)	(0.021)		
Age when promoted to Major	0.002*	0.001	0.003*	0.001	0.003**
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Individuals married = 1	0.014	0.012	0.015	0.010	0.018
	(0.009)	(0.010)	(0.013)	(0.009)	(0.014)
Average Race Similarity to 1LS and 2LS	-0.015	-0.008	0.021	-0.016	0.013
	(0.016)	(0.020)	(0.034)	(0.015)	(0.045)
Average Gender Similarity to 1LS and 2LS	-0.027	-0.040	-0.008	-0.031	0.012
	(0.021)	(0.027)	(0.027)	(0.022)	(0.053)
Years of Active Federal Military Service	0.003***	0.004***	0.003	0.004***	0.003
	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)
Officer has completed ILE or SSC = 1	0.126***	0.098***	0.098***	0.109***	0.066***
	(0.018)	(0.014)	(0.016)	(0.013)	(0.020)
Any Joint Professional Military Education = 1	0.003	0.002	0.009	-0.002	0.037
	(0.009)	(0.009)	(0.018)	(0.009)	(0.028)
Years from first rating officers to eligibility for LTC	-0.005	-0.003	-0.009	-0.003	-0.007
	(0.004)	(0.005)	(0.007)	(0.004)	(0.016)
Years as a supervisor of officers	0.025**	0.025***	0.018	0.021**	0.043
	(0.011)	(0.010)	(0.024)	(0.010)	(0.041)
Total number of supervisor evaluations (logged)	-0.002	-0.010	0.026	0.000	-0.023
	(0.015)	(0.013)	(0.038)	(0.015)	(0.064)
Branch similarity to supervisors	-0.040**	-0.041**	-0.046	-0.044*	-0.030
	(0.019)	(0.020)	(0.028)	(0.023)	(0.028)
Average Senior Rater Score (Scale:1-4)	0.184***	0.163***	0.239***	0.199***	0.102**
	(0.022)	(0.025)	(0.051)	(0.024)	(0.051)
Average Score of Rater and SR (Scale: 2-8)	0.181***	0.168***	0.207***	0.177***	0.203***
	(0.013)	(0.014)	(0.026)	(0.015)	(0.022)
Competitive Category: Operations = 1	0.011	0.017	-0.009	0.010	0.034
	(0.022)	(0.017)	(0.039)	(0.023)	(0.033)
Competitive Category: Operations Support = 1	0.018	0.026	-0.006	0.023	0.004
	(0.021)	(0.020)	(0.040)	(0.024)	(0.030)
Competitive Category: Force Sustainment = 1	0.025	0.036*	0.004	0.029	0.021
	(0.021)	(0.021)	(0.039)	(0.025)	(0.028)
Competitive Category: Army Medical Department = 1	0.076***	0.078***	0.097**	0.091***	0.072**
	(0.018)	(0.021)	(0.047)	(0.030)	(0.030)
Year Promoted to Major	-0.020***	-0.021***	-0.016**	-0.022***	-0.007
	(0.003)	(0.003)	(0.007)	(0.003)	(0.006)
Observations	10,595	7,390	3,205	8,808	1,787
Cluster	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.3: Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, by race and gender

VARIABLES	(1) Logit ITS dv_promo White Men	(2) Logit ITS dv_promo Non-White Men	(3) Logit ITS dv_promo White Women	(4) Logit ITS dv_promo Non-White Women
Post Policy: post == 1 = 1	0.051***	0.065**	0.046**	-0.014
	(0.012)	(0.025)	(0.023)	(0.036)
Age when promoted to Major	0.000	0.003	0.003*	0.005**
	(0.001)	(0.002)	(0.002)	(0.002)
Individuals married = 1	0.012	0.007	0.017	0.018
	(0.010)	(0.016)	(0.018)	(0.019)
Average Race Similarity to 1LS and 2LS	-0.018	0.050	0.044	-0.055
	(0.023)	(0.043)	(0.050)	(0.066)
Average Gender Similarity to 1LS and 2LS	-0.035	-0.021	0.027	-0.014
	(0.031)	(0.031)	(0.048)	(0.083)
Years of Active Federal Military Service	0.004***	0.004*	0.005	-0.000
	(0.001)	(0.002)	(0.004)	(0.003)
Officer has completed ILE or SSC = 1	0.114***	0.101***	0.048**	0.095**
	(0.015)	(0.018)	(0.023)	(0.039)
Any Joint Professional Military Education = 1	-0.003	0.005	0.101	0.016
	(0.009)	(0.021)	(0.064)	(0.037)
Years from first rating officers to eligibility for LTC	-0.004	0.000	0.009	-0.050***
	(0.005)	(0.007)	(0.011)	(0.017)
Years as a supervisor of officers	0.020*	0.019	0.097**	0.034
	(0.010)	(0.019)	(0.046)	(0.043)
Total number of evaluations as supervisor (logged)	0.002	0.003	-0.128***	0.122**
	(0.015)	(0.031)	(0.047)	(0.058)
Branch similarity to supervisors	-0.044*	-0.047	-0.050	-0.025
	(0.023)	(0.032)	(0.041)	(0.047)
Average Senior Rater Score (Scale:1-4)	0.182***	0.250***	0.025	0.179**
	(0.025)	(0.061)	(0.069)	(0.084)
Average Score of Rater and SR (Scale: 2-8)	0.165***	0.207***	0.178***	0.226***
	(0.014)	(0.032)	(0.038)	(0.041)
Competitive Category: Operations = 1	0.016	-0.015	0.038	0.014
	(0.017)	(0.044)	(0.032)	(0.045)
Competitive Category: Operations Support = 1	0.027	0.001	0.036	-0.025
	(0.019)	(0.045)	(0.039)	(0.036)
Competitive Category: Force Sustainment = 1	0.038*	0.001	0.037	0.015
	(0.021)	(0.044)	(0.035)	(0.033)
Competitive Category: Army Medical Department = 1	0.087***	0.081	0.041	0.108***
	(0.023)	(0.056)	(0.029)	(0.039)
Year Promoted to Major	-0.022***	-0.022***	-0.016**	0.005
	(0.003)	(0.007)	(0.007)	(0.009)
Observations	6,469	2,339	921	866
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

2.4.1 Robustness Check

For robustness, I reran the primary model with a linear probability model, to ensure that the model is stable to the primary finding and not sensitive to model specifications. The results of the robustness check can be found in Appendix A. The linear probability model provides comparable results of the same sign and magnitude as the logit model.

2.4.2 Sensitivity Check

I conducted two sets of sensitivity checks. First, I reran the model, focusing on the bottom 50% of the promotion cohorts based on the individual average performance scores at the rank of major. With approximately 86% of the sample being promoted, this approach draws attention to individuals close to the promotion decision threshold—where the partial blindness policy is likely to have the greatest impact on the subjective rankings of the promotion board. At the aggregate level, the only significant results are for all whites, all males, and specifically white males. Compared to the full panel, policy effects for white women are no longer conclusive.

For the second sensitivity check, I remove all promotions in 2021 to remove the transition year of the policy. I re-estimate Equation 2.1, reducing the promotion windows to two years, with post-policy promotions in 2022-2023, compared to 2019-2020 for the pre-policy group. The sensitivity check shows the same trends as the primary model and the robustness check when removing the 2021 promotion year from the results, supporting the primary findings. Descriptive statistics and results of this sensitivity check can be found in Appendix B.

2.4.3 Heterogeneity by Occupational Group

To extend the analysis, I stratify the sample by occupational workgroup, recognizing that groupings of occupational workgroups delineate the peer cohorts used for promotion consideration. Figure 2.2 shows stratified results for the core function employees. Here, I find an effect among white men ($\beta = 0.056, SE = 0.012$) and non-white men ($\beta = 0.087, SE = 0.031$) who see a statistically significant increase in promotion rates, respectively, due to the policy. Figure 2.3 shows stratified results for the intelligence and communications occupational workgroup. The primary finding from this graph is that the policy has a significant negative impact on non-white women ($\beta = -0.063, SE = 0.031$). In Figure 2.4, the results for the administrative, human resources, and logistics occupational workgroup show that the policy is only beneficial to white men, and inconclusive for the other demographic groups. Lastly, in Figure 2.5, in the medical administration and nursing occupational grouping, white men and non-white men see a boost in promotion from the policy. Regression results for these figures can be found in Appendix A, Tables A.2 to A.5.

When I rerun the model, focusing on the bottom 50% of the promotion cohorts among occupational groupings, among core function employees (Table B.4), only non-white men see an approximately 9 percent increase ($\beta = 0.089, SE = 0.042$). In the medical administration and nursing occupational group (Table B.7) non-white men still see an approximately 29 percent increase in promotion rates due to the policy ($\beta = 0.290, SE = 0.108$). Descriptive statistics and results tables for these models can be found in Appendix B. Interestingly, these results show that closer to the promotion cutoff point, the primary beneficiary of the policy continues to be men, but primarily non-white men.

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Core Function Employees

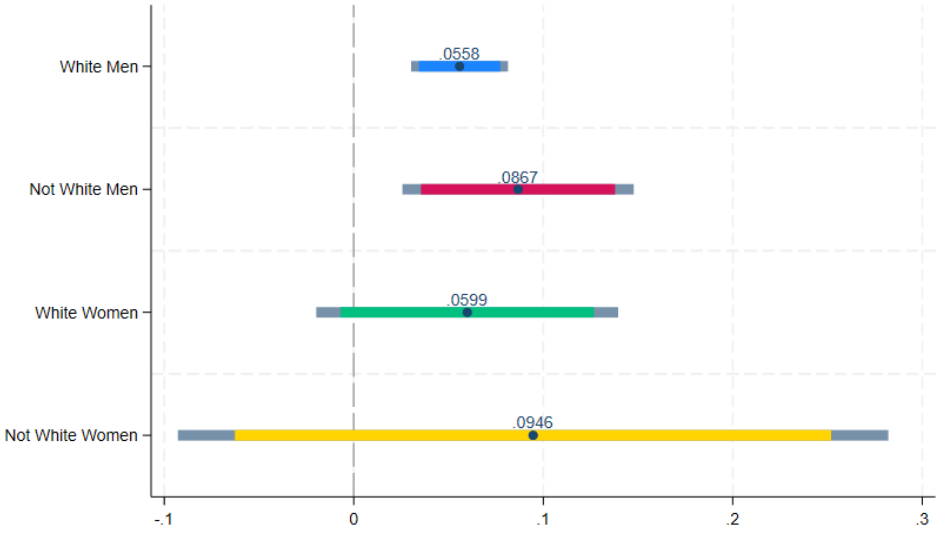

Figure 2.2: Primary Results of the SITS Model for Core Function Employees

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Intelligence and Communications Employees

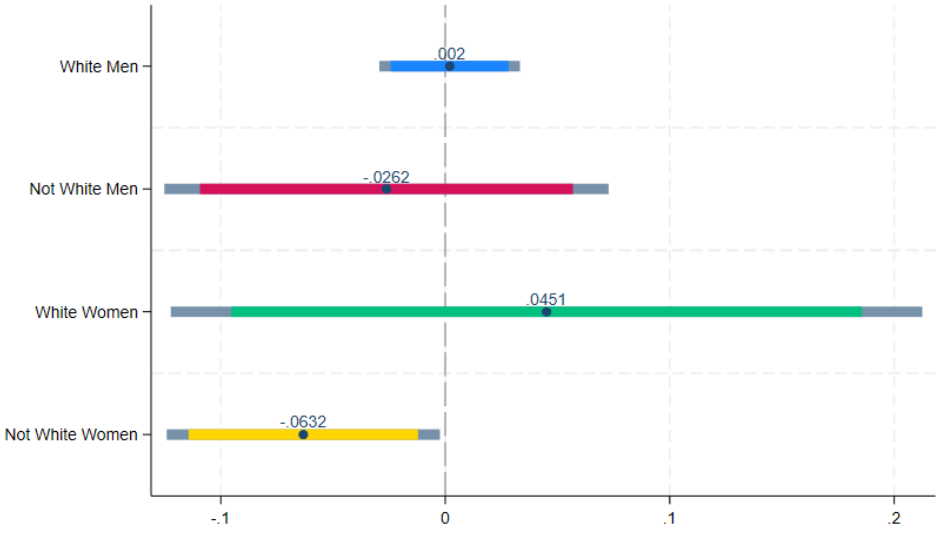

Figure 2.3: Primary Results of the SITS Model for Intelligence and Communications Employees

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Administrative, Human Resources, and Logistics Employees

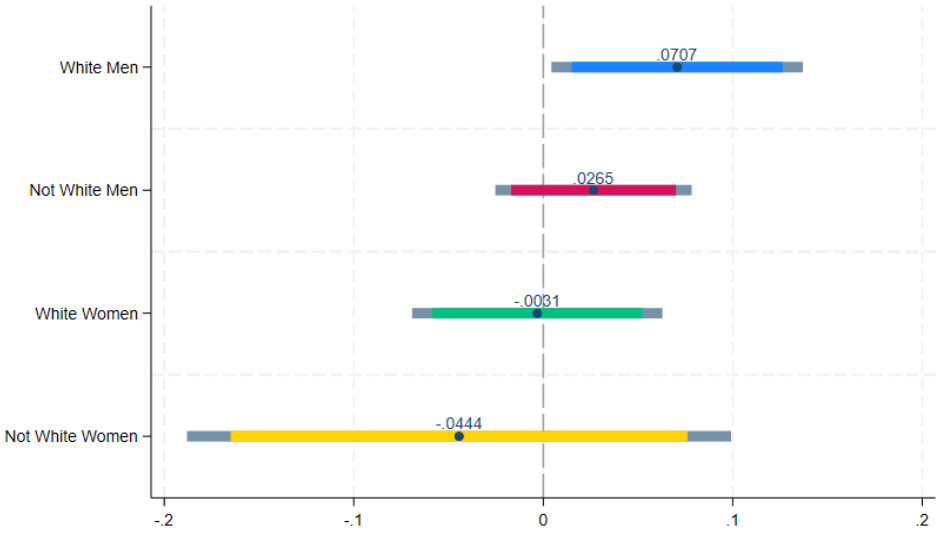

Figure 2.4: Primary Results of the SITS Model for Administrative, HR, and Logistics Employees

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Medical Administrative Employees and Nurses

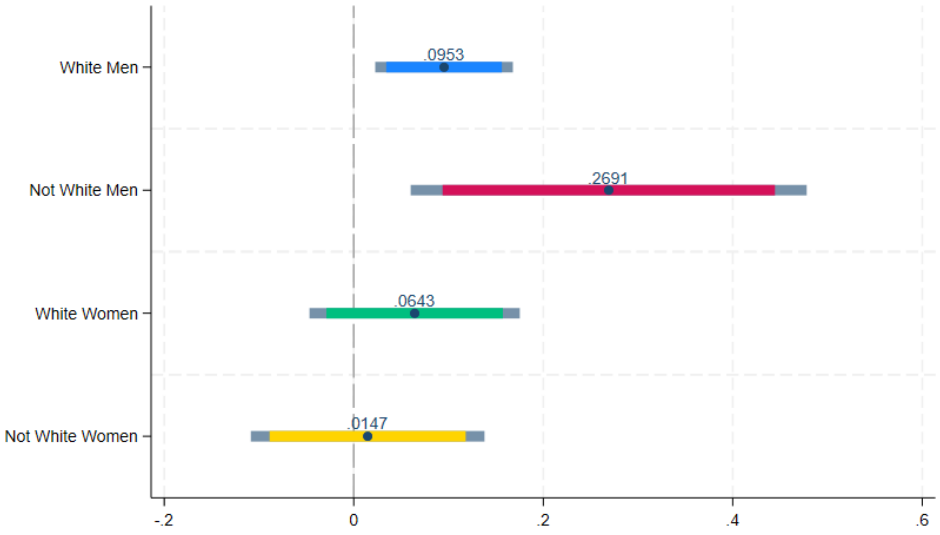

Figure 2.5: Primary Results of the SITS Model for Medical Administration and Nursing Employees

Further analysis of specific occupational groupings reveals several noteworthy findings. For example, among the intelligence and communications occupational group (Table B.5), the results show that as gender similarity with first and second-line supervisors increases probability of promotion decreases for white men ($\beta = 0.271, SE = 0.150$), non-white men ($\beta = -0.254, SE = 0.086$), and non-white women ($\beta = -1.384, SE = 0.408$). In the same occupational group, non-white women see a slight boost in their promotion chances when they have supervisors that are the same race ($\beta = 0.178, SE = 0.076$). Also, non-white women benefit when they increase the number of supervisors in their same occupational workgroup ($\beta = .197, SE = 0.076$). Although gender and racial impacts remain evident among the bottom 50% of promotion cohorts, occupational workgroup similarity with supervisors only shows an effect within this lower performance group. In the bottom 50% cohort, among core employee ($\beta = -0.132, SE = 0.042$) and administrative, human resources, and logistics occupational workgroups ($\beta = -0.148, SE = 0.072$), white men see a decrease in their promotion chances as their average occupational workgroup similarity with their supervisors increases.

2.4.4 Coarsened Exact Matching (CEM)

To further stress my model, I also run it with Coarsened Exact Matching. I match observations in the pre- and post-policy periods by race (white or not-white), gender (male or female), occupational specialization (Army military occupational specialty (MOS)), average performance scores while at the rank of major (broken into top, middle, and bottom thirds), and previous supervisor experience (yes or no). Of the original 10,595 observations, using a 1-to-1 match, I match 3,959 individuals in the pre- and post-policy period, respectively, in 580 of 1087 possible strata. Results for this method, shown in Figure 2.6 were similar in magnitude and direction to the primary

unmatched logit model. Detailed results can be found in Appendix A, Figure A.1 to Figure A.4.

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy

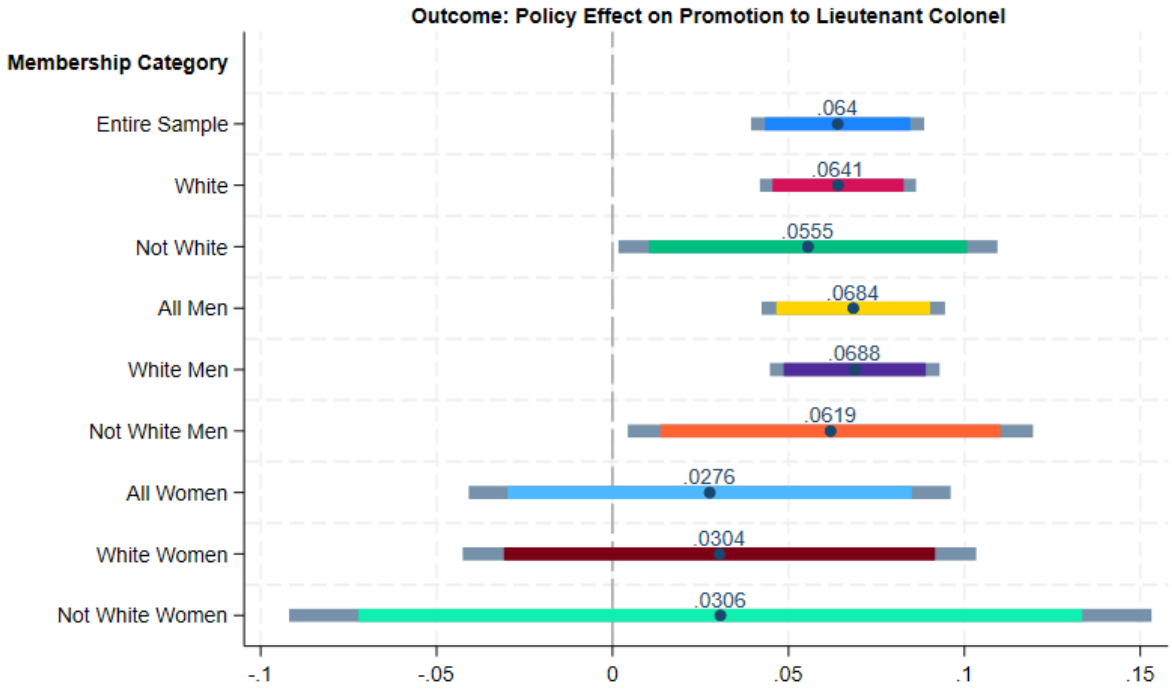

Figure 2.6: CEM Model - Full Sample

2.4.5 Relational Demography

There continues to be additional nuance that requires further exploration in regard to relational demography. The findings suggest that relational demography's impact on promotions to lieutenant colonel is across demographic groups at the aggregate level. Using relational demography, I assess the impact of employee's similarity to their supervisors for the duration of their time as a major. I compare similarity based on race, gender, and occupational workgroup by creating a similarity score. For each performance evaluation, the employee receives a 1 for each category (race, gender, and occupational workgroup, respectively) when they match their first or second

line supervisor (and a zero if they do not). Then, each score is summed and then divided by the total number of evaluations to create an average similarity value between 0 and 1. Finally, I determine the mean value for each category, and then assign the employee a binary value of one if their similarity score greater than or equal to the mean value for that category, or zero if their similarity value is less than the mean.

Table 2.4: Logit Regression Results for Relational Demography Model

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo Greater than mean racial similarity	(3) Logit ITS dv_promo Less than mean racial similarity	(4) Logit ITS dv_promo Greater than mean gender similarity	(5) Logit ITS dv_promo Less than mean gender similarity	(6) Logit ITS dv_promo Greater than mean branch similarity	(7) Logit ITS dv_promo Less than mean branch similarity
Post Policy: post == 1 = 1	0.051***	0.044***	0.059***	0.045***	0.060***	0.060***	0.045***
	(0.011)	(0.011)	(0.018)	(0.014)	(0.014)	(0.013)	(0.014)
All not white individuals = 1	-0.008	0.009	-0.013	0.023**	-0.029*	0.008	-0.014
	(0.010)	(0.021)	(0.011)	(0.010)	(0.015)	(0.014)	(0.013)
All females = 1	0.012	0.037**	-0.003	0.001	0.024	0.022	0.008
	(0.014)	(0.018)	(0.018)	(0.036)	(0.015)	(0.023)	(0.016)
Age when promoted to Major	0.002*	0.001	0.002	0.001	0.002	0.003**	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Individuals married = 1	0.014	0.021	0.008	0.011	0.017	0.027*	0.005
	(0.009)	(0.013)	(0.011)	(0.012)	(0.011)	(0.014)	(0.009)
Average Race Similarity to 1LS and 2LS	-0.015	0.019	-0.005	0.031	-0.051*	0.016	-0.032
	(0.016)	(0.028)	(0.037)	(0.020)	(0.030)	(0.021)	(0.024)
Average Gender Similarity to 1LS and 2LS	-0.027	-0.008	-0.034	-0.016	0.000	-0.021	-0.017
	(0.021)	(0.031)	(0.025)	(0.031)	(0.031)	(0.035)	(0.020)
Years of Active Federal Military Service	0.003***	0.003**	0.004**	0.003**	0.003*	0.004**	0.003**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Officer has completed ILE or SSC = 1	0.126***	0.135***	0.123***	0.143***	0.115***	0.113***	0.140***
	(0.018)	(0.036)	(0.020)	(0.028)	(0.019)	(0.022)	(0.027)
Any Joint Professional Military Education = 1	0.003	0.006	-0.001	-0.008	0.018	0.007	-0.000
	(0.009)	(0.010)	(0.013)	(0.011)	(0.011)	(0.015)	(0.011)
Years from first rating officers to eligibility for LTC	-0.005	-0.003	-0.005	0.002	-0.011**	-0.003	-0.007
	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)	(0.008)	(0.005)
Years as a supervisor of officers	0.025**	0.022	0.030*	0.024*	0.023	0.025**	0.030
	(0.011)	(0.013)	(0.017)	(0.014)	(0.014)	(0.012)	(0.019)
Total number of evaluations as supervisor	-0.002	-0.002	-0.007	-0.014	0.014	-0.006	0.001
	(0.015)	(0.019)	(0.021)	(0.018)	(0.015)	(0.022)	(0.022)
Branch similiarity with supervisors	-0.040**	-0.039*	-0.045**	-0.061**	0.003	-0.014	-0.045
	(0.019)	(0.022)	(0.021)	(0.025)	(0.021)	(0.033)	(0.030)
Average Senior Rater Score (Scale:1-4)	0.184***	0.194***	0.176***	0.220***	0.149***	0.217***	0.167***
	(0.022)	(0.035)	(0.027)	(0.037)	(0.027)	(0.043)	(0.022)
Average Score of Rater and SR (Scale: 2-8)	0.181***	0.174***	0.188***	0.204***	0.150***	0.171***	0.185***
	(0.013)	(0.019)	(0.015)	(0.020)	(0.016)	(0.023)	(0.015)
Competitive Category: Operations = 1	0.011	0.009	0.011	-0.002	0.035	-0.018	0.038
	(0.022)	(0.018)	(0.032)	(0.029)	(0.024)	(0.011)	(0.028)
Competitive Category: Operations Support = 1	0.018	0.023	0.011	0.001	0.038*	0.010	0.029
	(0.021)	(0.018)	(0.031)	(0.029)	(0.022)	(0.013)	(0.026)
Competitive Category: Force Sustainment = 1	0.025	0.035*	0.016	0.031	0.027	0.010	0.041
	(0.021)	(0.019)	(0.031)	(0.029)	(0.022)	(0.015)	(0.025)
Competitive Category: Army Medical Department = 1	0.076***	0.069***	0.081***	0.078***	0.072***	0.064***	0.079***
	(0.018)	(0.015)	(0.027)	(0.025)	(0.020)	(0.014)	(0.020)
Year Promoted to Major	-0.020***	-0.019***	-0.020***	-0.018***	-0.022***	-0.025***	-0.016***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.004)	(0.003)
Constant							
Observations	10,595	5,544	5,051	5,731	4,864	4,730	5,865
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The results from this model indicate that the intersection of race and gender significantly

influences advancement opportunities. On the whole, the policy was inconclusive to benefiting women for promotion. But, among women, as illustrated in Figure 2.7, having a greater number of supervisors of the same race, and even more notably, a higher number of supervisors of the same gender compared to their peers, is associated with a higher probability of promotion to lieutenant colonel. The results for the other levels of relational demography were inconclusive.[9]

Table 2.5: Logit Regression Results for Relational Demography Model, All Women

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo Greater than mean racial similarity	(3) Logit ITS dv_promo Less than mean racial similarity	(4) Logit ITS dv_promo Greater than mean gender similarity	(5) Logit ITS dv_promo Less than mean gender similarity	(6) Logit ITS dv_promo Greater than mean branch similarity	(7) Logit ITS dv_promo Less than mean branch similarity
Post Policy: post == 1 = 1	0.024	0.072**	-0.005	0.229***	0.013	0.023	0.025
	(0.020)	(0.035)	(0.033)	(0.043)	(0.021)	(0.030)	(0.032)
Age when promoted to Major	0.003**	0.004	0.003	0.004**	0.004**	0.005***	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Individuals married = 1	0.019	0.039*	0.006	0.234***	0.012	0.048**	-0.005
	(0.015)	(0.021)	(0.017)	(0.060)	(0.015)	(0.023)	(0.015)
Years of Active Federal Military Service	0.003	0.006	0.001	0.006**	0.003	0.001	0.005
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Officer has completed ILE or SSC = 1	0.082***	0.035	0.115**	0.118**	0.092***	0.079**	0.081
	(0.029)	(0.040)	(0.050)	(0.059)	(0.031)	(0.034)	(0.060)
Any Joint Professional Military Education = 1	0.033	0.010	0.039	-0.105***	0.039*	0.065*	0.005
	(0.024)	(0.039)	(0.034)	(0.027)	(0.023)	(0.038)	(0.032)
Years from first rating officers to eligibility for LTC	-0.008	0.021	-0.009	1.926***	-0.023*	0.008	-0.597***
	(0.016)	(0.029)	(0.019)	(0.491)	(0.013)	(0.024)	(0.055)
Years as a supervisor of officers	0.044	-0.001	0.089	0.178	0.075	0.022	2.439***
	(0.038)	(0.022)	(0.056)	(0.176)	(0.051)	(0.042)	(0.252)
Total number of evaluations as supervisor (logged)	-0.024	-0.007	-0.079	-2.167***	-0.015	-0.019	-0.036
	(0.063)	(0.066)	(0.073)	(0.356)	(0.043)	(0.090)	(0.044)
Average Senior Rater Score (Scale:1-4)	0.100**	0.117	0.088	-0.404***	0.111**	0.192**	0.053
	(0.049)	(0.077)	(0.068)	(0.129)	(0.048)	(0.084)	(0.058)
Average Score of Rater and SR (Scale: 2-8)	0.206***	0.136***	0.247***	0.481***	0.194***	0.153***	0.228***
	(0.022)	(0.040)	(0.030)	(0.073)	(0.022)	(0.046)	(0.026)
Competitive Category: Operations = 1	0.030	0.009	0.041		0.022	0.036	0.025
	(0.027)	(0.035)	(0.031)		(0.029)	(0.038)	(0.031)
Competitive Category: Operations Support = 1	0.001	0.019	0.007		-0.005	0.002	0.002
	(0.027)	(0.037)	(0.031)		(0.029)	(0.040)	(0.031)
Competitive Category: Force Sustainment = 1	0.019	-0.011	0.041		0.011	0.020	0.016
	(0.025)	(0.036)	(0.029)		(0.027)	(0.037)	(0.028)
Competitive Category: Army Medical Department = 1	0.065**	0.011	0.105***	-0.563***	0.062**	0.054	0.084***
	(0.026)	(0.035)	(0.029)	(0.188)	(0.027)	(0.041)	(0.023)
Year Promoted to Major	-0.007	-0.024***	0.004	-0.017	-0.007	-0.009	-0.007
	(0.006)	(0.009)	(0.008)	(0.015)	(0.006)	(0.010)	(0.009)
Observations	1,787	713	1,074	59	1,726	858	929
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

However, among non-white employees the only significant demographic intersection that helped improve promotion chances was having an above mean number of supervisors of

[9]I had an initial suspicion that nurses may be skewing the sample, as they made up half of the women in the sample. I ran additional tests for nurses and all women who were not nurses, respectively. Results were similar in direction and magnitude together and separately for these groups of females.

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Women

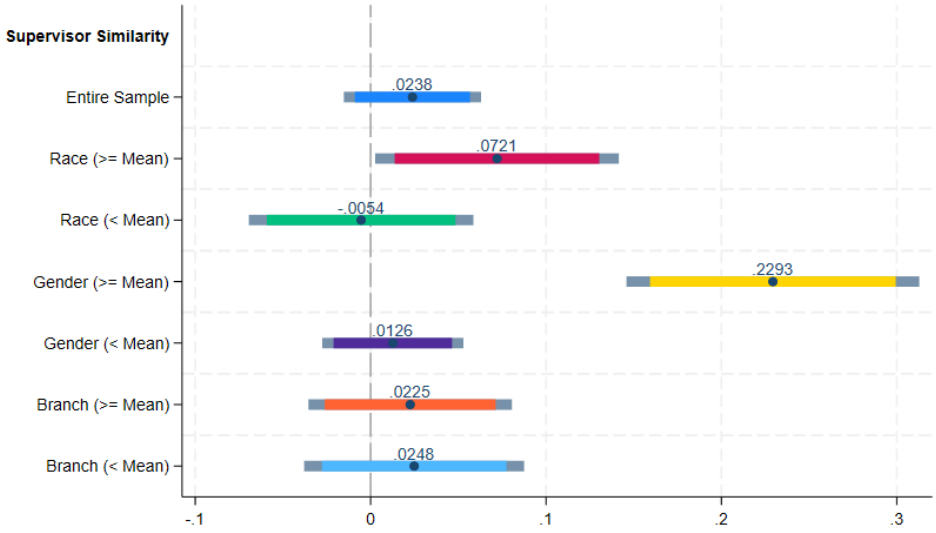

Figure 2.7: Primary Results of the Relational Demography Model for Female Employees

the same occupational workgroup (branch), as shown in Figure 2.8. Results for the other demographic intersections was inconclusive. Additional tables for relational demography can be found in Appendix A. Due to sample size and statistical power issues, I was unable to run an acceptable model specifically for non-white women.

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Non-Whites

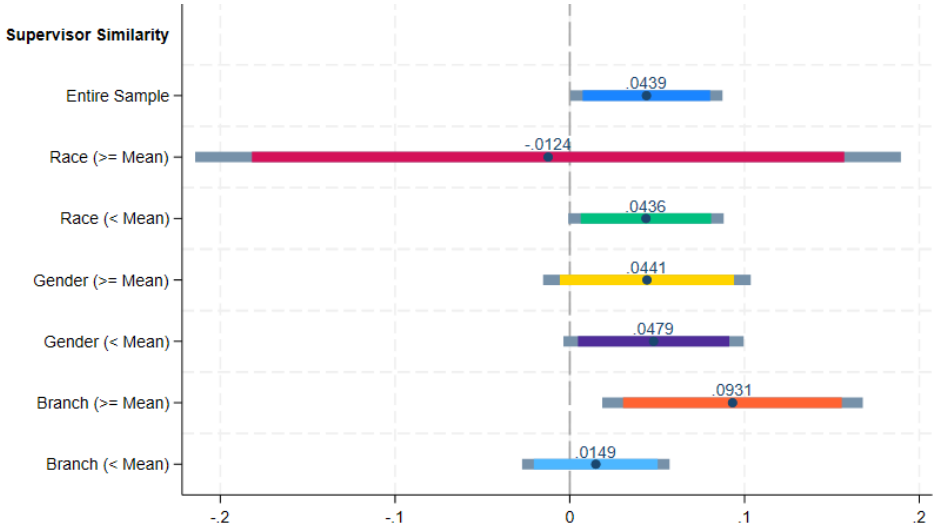

Figure 2.8: Primary Results of the Relational Demography Model for Non-White Employees

Table 2.6: Logit Regression Results for Relational Demography Model, Non-Whites

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo Greater than mean racial similarity	(3) Logit ITS dv_promo Less than mean racial similarity	(4) Logit ITS dv_promo Greater than mean gender similarity	(5) Logit ITS dv_promo Less than mean gender similarity	(6) Logit ITS dv_promo Greater than mean branch similarity	(7) Logit ITS dv_promo Less than mean branch similarity
Post Policy: post == 1 = 1	0.044**	-0.012	0.044*	0.044	0.048*	0.093**	0.015
	(0.022)	(0.103)	(0.023)	(0.030)	(0.026)	(0.038)	(0.022)
Age when promoted to Major	0.003**	-0.006	0.004**	0.002	0.004**	0.006**	0.001
	(0.002)	(0.008)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Individuals married = 1	0.010	0.058	0.007	-0.007	0.019	0.023	0.004
	(0.012)	(0.077)	(0.011)	(0.021)	(0.014)	(0.021)	(0.012)
Years of Active Federal Military Service	0.003	0.020***	0.002	0.005**	0.001	0.004*	0.002
	(0.002)	(0.007)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Officer has completed ILE or SSC = 1	0.122***	0.035	0.131***	0.112***	0.120***	0.126***	0.127***
	(0.023)	(0.072)	(0.023)	(0.031)	(0.033)	(0.035)	(0.033)
Any Joint Professional Military Education = 1	0.008	-0.035	0.011	-0.037	0.046**	-0.022	0.025
	(0.018)	(0.076)	(0.019)	(0.030)	(0.020)	(0.036)	(0.018)
Years from first rating officers to eligibility for LTC	-0.008	-0.280***	-0.006	-0.001	-0.014	-0.020	-0.005
	(0.007)	(0.073)	(0.008)	(0.006)	(0.014)	(0.020)	(0.007)
Years as a supervisor of officers	0.018	-0.365**	0.033**	-0.000	0.023*	0.018	0.037*
	(0.023)	(0.149)	(0.013)	(0.045)	(0.012)	(0.025)	(0.021)
Total number of evaluations as a supervisor (logged)	0.023	1.239***	-0.004	0.036	0.026	0.051	-0.014
	(0.038)	(0.339)	(0.025)	(0.064)	(0.040)	(0.066)	(0.023)
Average Senior Rater Score (Scale:1-4)	0.234***	0.397**	0.231***	0.215***	0.266***	0.261***	0.228***
	(0.050)	(0.198)	(0.051)	(0.067)	(0.066)	(0.062)	(0.060)
Average Score of Rater and SR (Scale: 2-8)	0.211***	0.249**	0.209***	0.253***	0.160***	0.184***	0.215***
	(0.026)	(0.102)	(0.027)	(0.036)	(0.035)	(0.032)	(0.031)
Competitive Category: Operations = 1	-0.009	0.122**	-0.019	-0.030	0.011	-0.085**	0.042
	(0.036)	(0.062)	(0.038)	(0.049)	(0.058)	(0.040)	(0.037)
Competitive Category: Operations Support = 1	-0.001	0.178***	-0.012	-0.026	0.005	-0.055	0.029
	(0.036)	(0.054)	(0.038)	(0.050)	(0.058)	(0.040)	(0.038)
Competitive Category: Force Sustainment = 1	0.012	0.203***	-0.000	0.018	0.001	-0.048	0.044
	(0.033)	(0.051)	(0.035)	(0.046)	(0.057)	(0.042)	(0.036)
Competitive Category: Army Medical Department = 1	0.086**	0.310***	0.072**	0.079*	0.080	0.034	0.123***
	(0.034)	(0.050)	(0.036)	(0.042)	(0.051)	(0.046)	(0.028)
Year Promoted to Major	-0.015**	-0.001	-0.016**	-0.016*	-0.016**	-0.032***	-0.005
	(0.006)	(0.026)	(0.006)	(0.008)	(0.007)	(0.008)	(0.006)
Observations	3,205	139	3,066	1,517	1,688	1,325	1,880
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

2.5 Discussion

I examined how the implementation of a partial blindness policy, in a large governmental organization, affected promotion at an upper management level. The sample was two cohorts of U.S. Army majors that sought promotion in the three years before and after the policy's implementation. Using a stratified interrupted time series (SITS) logit regression to evaluate the change in promotions rates within demographic and occupational group membership, I fail to find an effect for underrepresented minority groups, which offers evidence supportive of hypothesis 1 (H1). For hypothesis 2 (H2), I find that relational demography operates in nuanced ways, reflecting a complex interplay between intersectional identity and occupational context. The observed heterogeneity in policy effects reveals how structural and interpersonal dynamics jointly shape promotion outcomes, especially in hierarchical systems where ambiguity and discretion amplify bias.

Similarity based biases appear more influential under conditions of greater subjectivity, such as among borderline candidates, and within occupational groups where conformity norms or identity signaling may play a stronger evaluative role. For white men, dominant group status within most occupational hierarchies may allow them to benefit most from partial blindness policies, which obscure overt markers of bias while preserving underlying structural advantages. Non-white men may experience modest gains from reduced racial visibility, yet these benefits are constrained by outsider status in traditionally white-dominated fields. The inconclusive or adverse results for white and non-white women highlight the compounded effects of gender and racial bias that partial blindness fails to mitigate. Taken together, these patterns suggest that promotion decisions may not be purely meritocratic, as systemic inequities and relational dynamics

continue to shape outcomes. Thus, the limitations of one-size-fits-all policies become apparent, pointing to the need for more tailored approaches that consider the social architecture of evaluation and ensure individuals are assessed on merit rather than accumulated disadvantage.

2.5.1 Mechanisms

To further interpret these findings, it may be necessary to trace the promotion trajectories back to the very beginning of officers' careers. Although the current analysis is structured at the individual-year level, the competitive nature of the up-or-out system suggests that potentially even small timing advantages, measured in months rather than years, can accumulate over time. These incremental gains may lead to earlier access to key developmental assignments and supervisory roles, providing certain officers with critical leadership experience ahead of their peers. As these advantages compound, they can create pathways to accelerated career progression that are not easily offset by later evaluations. To explore this mechanism more deeply, future research should examine officer placements and supervisory assignments at the quarterly level to capture the finer gradations of career advancement. This more granular analysis would allow for a closer investigation into how intersectionality, across both individual characteristics (e.g., race, gender) and professional attributes (e.g., branch, functional area), shapes access to critical career-building opportunities in the earliest stages of service.

Another potential mechanism that may influence promotion outcomes under partial blindness policies is the role of networking and sponsorship. In hierarchical organizations like the Army, professional networks serve as critical channels for accessing career-enhancing opportunities, including leadership roles, special assignments, and mentorship. Officers who establish strong networks early in their careers are more likely to receive sponsorship from senior leaders,

which can translate into more favorable evaluations and accelerated promotion trajectories. This sponsorship is often facilitated through informal relationships built during key assignments or shared professional experiences.

A critical mechanism that may further explain disparities in promotion outcomes is the role of intersectionality, particularly when examined through matched comparison techniques such as Coarsened Exact Matching (CEM). Intersectionality posits that overlapping social identities, such as race and gender, create unique experiences of advantage or disadvantage within organizational processes. Traditional evaluation methods often overlook these compounded effects, treating demographic characteristics as independent rather than interlinked. While this study conducted an initial CEM analysis on race, gender, occupational specialty, performance score tercile, and a binary indicator for previous supervisory experience, a more detailed range of covariates including years of service, separate similarity variables for first- and second-line supervisors, and when the employee achieved their first supervisory position, would allow for a more refined assessment of how intersectional identities influence promotion likelihood under partial blindness policies. This approach would facilitate a clearer understanding of whether the policy disrupts existing biases or simply shifts them across demographic intersections.

2.5.2 Theoretical Contributions

Most research on blindness policies focuses on hiring decisions, examining how concealing demographic information can reduce biases against under-represented groups. While this literature shows that blindness policies can improve diversity in initial hiring, there is limited understanding of how these policies affect internal promotions. By studying the promotion process rather than entry-level hiring, I expand blindness research to an intra-organizational context, where

decision-makers are more familiar with candidates and institutional biases may play a greater role. This shift addresses a critical gap by investigating how blindness policies impact meritocratic decision-making in mid-career promotions, offering insights into whether these policies have similar effects in contexts influenced by hierarchical relationships and prior knowledge of candidates.

Second, a significant theoretical contribution of this study lies in examining the limitations of partial blindness policies. Unlike full blindness, which conceals all identifying information, partial blindness leaves certain demographic markers visible, in this case names and pronouns. This study reveals that partial blindness, in this instance, did not achieve the stated goal of increasing diversity, as biases persisted despite the partial concealment of identity. These findings challenge the assumption that any level of blindness will yield equitable outcomes, suggesting instead that partial blindness may allow residual biases to operate. This nuanced insight extends the literature by highlighting how varying degrees of information concealment impact the effectiveness of blindness policies, particularly in promotions. It suggests that partial blindness may offer only limited protection against bias and may, in fact, require complementary strategies to achieve meaningful processes that are genuinely meritocratic.

2.5.3 Practical Implications

Assessing middle- to upper-level managers, whose responsibilities include leadership, prioritization, and delegation, is more complex than evaluating roles with tangible and easily measurable outcomes. Additionally, in settings where coordination and synchronization are essential, it is impractical to indefinitely conceal employees' group membership characteristics from supervisors.

This study's findings show that promotion outcomes for individuals other than white men,

by race or gender, do not significantly improve under this partial blindness policy. This suggests that partial blindness alone may be insufficient to create equitable opportunities for advancement. In some cases, partial blindness might even exacerbate inequities by masking underlying structural biases without fully concealing identity, leading decision-makers to rely on inferred or remaining cues in ways that reinforce existing disparities. To counter this issue, organizations may benefit from complementary diversity initiatives that directly address structural biases to foster more inclusive promotion practices (Baker et al., 2021).

Another issue to consider for blindness policies is the logistical feasibility of implementing them due to the labor and process costs associated with changing entire human resource systems. In a study from the Canadian government (Public Service Commission of Canada, 2018), it took 15-20 minutes to manually redact each of the 2,226 resumes across 27 government hiring processes that were piloted, equivalent to 14 workweeks of additional labor for one person. Additionally, only up to 100 applications per position were anonymized. Anonymizing the entire process for these 27 positions would have required even more effort.

2.5.4 Policy Implications

These findings suggest that organizations and regulatory bodies should carefully evaluate whether partial blindness policies alone are sufficient to promote gender and race neutral advancement. Policies that obscure only selected demographic indicators may not achieve equitable meritocratic outcomes in subjective decisions like promotions. Moreover, as Apfelbaum et al. (2012) highlight, there is a potential trap: by making race or group identity salient, some individuals may shift to a procedural interpretation of blindness, advocating for equal treatment without accounting for pre-existing inequalities. Policymakers may need to consider whether policies should

instead enhance the visibility of group membership to address historical inequities directly. By explicitly recognizing and compensating for the challenges faced by under-represented groups, policies could better align with diversity goals and contribute to meaningful shifts in representation. This approach would ground diversity efforts in structural inclusivity rather than relying solely on concealment to counteract bias.

2.5.5 Limitations and Future Research

This study has several contextual limitations that should be considered. First, it provides only a limited snapshot of the effects of partial blindness policies, as the policy was recently implemented, allowing only for an initial analysis of short-term outcomes. Additionally, recent years have seen unusually high Army promotion rates (Lytell et al., 2023), which may influence the generalizability of the findings. The analysis also focuses on broad, cohort-based promotions rather than advancement to specific, individually selected positions, which may yield different dynamics. Furthermore, while the study examines the effects of partial blindness policies on promotion, it cannot fully isolate the impact of these policies from the larger organizational context, where various other factors influence promotion outcomes.

Other limitations to this study are that it relies on performance appraisals as a primary measure of promotion potential, even though implicit biases may still influence evaluations despite obscuring demographic information. While meritocratic ideals underlie these evaluations, other research suggests that such ideals often do not manifest in practice, as underrepresented minorities and women continue to face significant obstacles to promotion (Castilla and Benard, 2010; González Ramos et al., 2015).

Early career experiences, such as exposure to harassment or demeaning comments, may

shape long-term aspirations and thus influence promotion outcomes (Azmat et al., 2020). Research suggests that career privileges, rather than exclusionary practices, often determine who ascends to higher ranks, reinforcing existing disparities (Wildman et al., 1996). Moreover, as Peña et al. (2017) point out, the greatest barriers to inclusivity often emerge within the first five years of a career. Additional factors such as access to role models, career preferences, training opportunities, and retention decisions collectively influence diversity in senior leadership roles across race, ethnicity, and gender (Butler and Denton, 2021). This study's post-hoc analysis of Army promotions at the promotion board level cannot fully account for other process biases, such as mentorship access, career "anointings," and opportunities that reflect institutional privilege. Finally, structural factors, including entry source, qualifications, job assignments, and promotion criteria, also play critical roles in shaping career progression (Asch et al., 2012), underscoring that factors beyond individual merit significantly impact promotion outcomes.

An additional dimension that warrants further investigation is the role of supervisory experience across an officer's career. In the white-dominated core employee group, supervisor experience appears to enhance promotion prospects for non-white men, potentially serving as a legitimizing signal of leadership ability that offsets outsider status. However, this benefit is not universal. In occupational groups with strong technical identities, such as nursing and medical administration, greater supervisory experience is associated with a decrease in promotion likelihood for non-white men and white women. For individuals in these roles, heavy supervisory duties may be interpreted as a departure from core technical competence, inadvertently undermining perceptions of subject-matter expertise. In contrast, non-white women in these same technical fields appear to benefit from increased supervisory experience, perhaps because it signals both leadership and boundary-crossing ability in a space where they are doubly underrepresented.

These divergent effects suggest that the relationship between supervision and promotion is contingent not only on demographic identity but also on occupational norms and evaluative schemas. As such, future research should more closely examine how supervisory experience is interpreted across occupational contexts and identity intersections, particularly in fields where leadership and technical credibility are perceived as mutually exclusive.

2.6 Conclusion

This study contributes to the ongoing conversation about the effectiveness of blindness policies by exploring how partial blindness impacts promotion outcomes within a large organizational context. While full blindness policies have shown success in reducing biases in hiring, the results of this study indicate that partial blindness policies do not achieve the desired impartial meritocratic outcomes through internal promotions. This result highlights the need for organizations to consider additional identity-conscious strategies alongside partial blindness policies to address systemic biases.

For human resource professionals and policymakers, these findings suggest that partial blindness policies, while a step forward, may not be enough to fully address disparities in promotion outcomes. Organizations may need to implement more comprehensive approaches that tackle the structural and procedural barriers faced by the less represented groups in their organization. Future research could extend these findings by examining how similar policies affect promotion in other industries, or by testing whether combining partial blindness with targeted management initiatives leads to more meritocratic outcomes.

Chapter 3

DUAL PERSPECTIVES: EMPLOYEE PERFORMANCE AND SUSTAINABILITY IN EMPLOYEE-SUPERVISOR DYADS

3.1 Introduction

Organizations often face a trade-off between short-term productivity and long-term resource depletion (Maslach, 1982; Barnes et al., 2023). Supervisors must evaluate employees' performance in contexts where absenteeism, stress, and wellness behaviors are increasingly common. Attribution theory suggests that supervisors often default to internal explanations for employee absences or struggles, leading to harsher evaluations when the true causes are unknown or invisible (Swift et al., 2013). Yet emerging work on human sustainability leadership challenges traditional appraisal models, arguing that supervisors should prioritize not only output but also the long-term psychological, emotional, and physical health of their employees (Luthans and Church, 2002; Barnes et al., 2023). This study investigates how U.S. Army supervisors' pursuit of behavioral health (BH) care influences their evaluations of subordinates. By integrating insights from attribution theory and human sustainability leadership, I explore whether supervisors' personal experiences with wellness support mitigate negative biases in performance evaluations within an

organization that publicly promotes self-care and well-being (Acosta et al., 2014).

To investigate these dynamics, I leverage a unique and rich dataset of active-duty U.S. Army officers from 2010 to 2024, focusing on employee-supervisor dyads. I examine how supervisors' own behavioral health (BH) seeking behaviors influence their evaluations of subordinate employees. I link employee annual performance evaluations, personnel records, and outpatient mental health appointment records using unique individual identifiers. The Army's hierarchical and high-stress environment provides an ideal context to study attribution biases and human sustainability leadership, as it formally encourages wellness behaviors while simultaneously maintaining a competitive, output-driven appraisal system (DiBenigno, 2020; Barnes et al., 2023). This setting enables a strong test of whether supervisors' personal experiences with BH reshape their attribution processes and performance assessments.

This study makes several contributions to the management literature. Theoretically, it extends employee-supervisor research by examining how supervisors' personal wellness behaviors shape their performance evaluation decisions, connecting employee sustainability literature with performance management research (Maslach, 1982). The work advances leadership theory by incorporating wellness perspectives into evaluation frameworks (McGuffin et al., 2021). Practically, the findings provide organizations with insights for balancing performance demands and workforce sustainability in high-stress environments, which can help organizations design policies that encourage supervisor self-care without compromising fair and accurate employee evaluations (Acosta et al., 2014). By identifying how supervisors' wellness behaviors affect performance ratings, this research can support broader efforts to build healthier, more sustainable work environments, especially in high-stress, high-stakes organizations like the military (Barnes and Wagner, 2023).

3.2 Literature Review

This study builds on several core theoretical pillars. Central to workplace performance evaluations is the relationship between supervisors and employees (Yang et al., 2024; Tsui and O'Reilly III, 1989). Attribution theory suggests that supervisors often default to internal explanations (e.g., the employee is a poor worker) for absenteeism when the underlying causes are unknown, overlooking situational factors (Corrigan, 2000). Related research on workplace absenteeism shows that unexplained absences systematically bias supervisors' perceptions of employee commitment and reliability (Kreiner et al., 2022). Emerging work on human sustainability leadership challenges traditional appraisal models by emphasizing the supervisor's role in fostering long-term employee well-being alongside performance (Barnes and Wagner, 2023). Together, these literatures suggest that supervisors' own experiences with behavioral health seeking may moderate attributional biases and promote more supportive evaluations of employees. The following sections develop each of these foundations in turn.

3.2.1 Employee-Supervisor relationships and why they are important

The relationship between employees and supervisors is a critical factor in organizational behavior, influencing employee performance, job satisfaction, and overall well-being (DiBenigno, 2020). This relationship is particularly important in high-stress environments, such as the military, where job demands can lead to burnout and disengagement (Schaufeli et al., 2009). Supervisors play a crucial role in employee success by evaluating performance and providing feedback. In this context, the relationship between employees and supervisors becomes increasingly critical, as supervisors have the power to shape employees' perceptions of their own performance and

potential for advancement (DiBenigno and Kerrissey, 2020). Research shows that well-meaning managers can demonstrate bias in performance evaluations, despite organizational claims of meritocracy (Castilla and Benard, 2010).

Furthermore, Castilla (2008) demonstrated that with comparable evaluations, a second stage of bias, known as performance-reward bias, can emerge. Despite identical performance scores, women and minorities often receive lower compensation than their white male counterparts due to differences in gender, race, or ethnicity. This bias can be exacerbated by the competitive nature of the workplace, where employees are often pitted against each other for promotions and recognition (Castilla and Ranganathan, 2020).

3.2.2 Workplace Absenteeism, Attribution, and Supervisor Perceptions

Absenteeism, particularly when frequent or unexplained, often triggers negative supervisor perceptions. Research consistently links absenteeism to job stressors and lower performance ratings. For example, Parker and Kulik (1995) found that employees experiencing burnout were absent more often and received lower performance evaluations, while Richmond et al. (2014) showed that absenteeism predicted poorer ratings even after controlling for other factors. These patterns suggest that time away from work, regardless of cause, can negatively influence appraisals, exacerbating burnout and turnover risks (Barnes et al., 2023).

Attribution theory helps explain why absenteeism may so strongly influence supervisor judgments. Managers often commit correspondence bias (fundamental attribution error) by overweighting internal causes of behavior and overlooking situational factors (Swift et al., 2013). For example, supervisors may assume that an absent or underperforming employee is lazy or uncommitted rather than considering external pressures such as illness. Judge and Martocchio (1995)

demonstrated that supervisors who attributed absences to internal causes imposed harsher disciplinary decisions than those who recognized external explanations. Because supervisors often lack information about the true cause of an absence, particularly for employee well-being concerns, they may default to negative assumptions. Fear of stigmatization further discourages employees from disclosing mental health challenges, reinforcing these attributional biases (Mendel et al., 2013; Acosta et al., 2014; Kreiner et al., 2022). As a result, employees with undisclosed health issues may receive unfairly negative performance appraisals.

3.2.3 Empirical Links Between Manager Well-Being and Employee Evaluations

Research directly tying a supervisor's own mental health status to their rating leniency or stringency is scarce, but several lines of evidence hint at a connection. The most direct study of rater psychology by Parker and Kulik (1995) concerned burnout of employees rather than supervisors, yet it showed burnout correlated with worse supervisor ratings. This finding implies that stress-related phenomena link to appraisal outcomes. Another relevant piece comes from managers' own feedback. Castilla and Ranganathan (2020) found that supervisors draw from their own experiences as employees to shape their perceptions of what is important in evaluations. Crucially, recent evidence confirms that a supervisor's attitudes and well-being shape how they evaluate subordinates. The Swedish study by (Hultqvist et al., 2024) showed that a small negativity in attitude (only 2% variance) caused lower ratings of employees with health problems. Similarly, Mendel et al. (2013) found that giving identical work scenarios different labels (e.g., "personal issue" versus "behavioral problem"), led to supervisors rating employees differently. This suggests that the framing of an employee's situation interacts with supervisor perceptions, potentially exacerbating bias in evaluations.

While these examples provided do not test the supervisor's own mental health directly, they underscore that supervisor psychology (attitudes, mood, stress) do influence appraisal ratings. A frustrated or unsympathetic supervisor, whether frustrated by their workload, experiencing burnout, or dealing with undisclosed health issues, is likely to exhibit some or all of these documented biases. Overall, the empirical literature suggests that both subordinate factors (absences, known health problems) and supervisor factors (attitudes about mental illness, possibly their own stress) affect performance ratings.

What remains largely unexplored is precisely how a supervisor's personal pursuit of well-being (e.g. behavioral health history) impacts their evaluations of others. The expectation from theory is that a manager who has experienced mental health struggles may either be more empathetic or potentially more sensitive to related triggers. At present, the literature reveals a complex interplay in which high levels of stress or stigma in supervisors correlate with more negative evaluations of employee performance (Parker and Kulik, 1995). This pattern points to the need for further research on supervisor well-being as a contextual factor in appraisal accuracy and fairness.

3.2.4 Balancing Attribution Bias with Human Sustainability Leadership

While attribution theory suggests supervisors often penalize employees for unexplained absences or perceived lack of commitment, human sustainability leadership offers an alternative framework. Effective supervisors are responsible not only for immediate output but also for the long-term psychological, emotional, and physical well-being of their employees (Barnes and Wagner, 2023), reframing them as stewards of human resources rather than solely evaluators of productivity.

This dual responsibility creates a managerial paradox in which supervisors must assess performance, often shaped by absenteeism or stress, while also fostering a work environment that supports well-being. Traditional appraisal models emphasize efficiency and attendance, whereas sustainable leadership prioritizes care, empathy, and resilience-building (Barnes et al., 2023).

Supervisors with personal experience seeking behavioral health care may be more attuned to these hidden burdens, adopting contextual rather than dispositional attributions for employee wellness behaviors. Leaders who model vulnerability are seen as more authentic and supportive, fostering trust and improving long-term performance (Lemoine et al., 2019). Yet, empathetic supervisors operate within systems that prioritize measurable outputs (DiBenigno, 2018). This tension between personal empathy and organizational pressure may shape how supervisors evaluate employees who seek behavioral health support.

The Restricted Employee Sustainability Theory (REST) theorizes that employees' ability to sustain high performance depends on leadership behaviors that support regenerative capacity, such as encouraging wellness (Barnes and Wagner, 2023). REST highlights how leadership behaviors that promote sustainability, like work-life balance and mental health support, enhance both individual and organizational outcomes (Barnes et al., 2023; Hammer et al., 2024). Alignment between supervisor and employee values can strengthen resource sustainability (Fleming, 2024; Geibel et al., 2022), yet evidence remains limited on how these dynamics manifest in hierarchical relationships (Campbell et al., 2023; Hu et al., 2023).

3.2.5 Hypotheses Development

Building on this theoretical framework, I propose that supervisors' behavioral health-seeking behavior influences how they evaluate their subordinates, particularly when there is alignment

in personality traits linked to empathy, resilience, and openness to seeking support. Such trait similarity may enhance psychological safety and trust within the dyad, as supervisors may be more inclined to interpret subordinates' behaviors through a lens of understanding and empathy (Kristof-Brown et al., 2005; van Vianen et al., 2011). This alignment in dispositions may mitigate attributional biases related to absenteeism or perceived lack of commitment, fostering a more supportive evaluative process.

In contrast, when employees seek behavioral health care but their supervisors do not, the lack of alignment in attitudes toward behavioral health, specifically, or wellness-seeking behaviors, more broadly, may contribute to misunderstandings and bias in performance evaluations (Acosta et al., 2014; Kreiner et al., 2022). Supervisors who do not prioritize or understand behavioral health support, or wellness seeking, may be more inclined to attribute wellness behaviors to personal deficiencies rather than contextual challenges, reinforcing stigma and leading to lower evaluations. This gap in perception may create implicit bias, where supervisors undervalue the legitimacy of behavioral health needs, affecting their assessment of commitment and reliability.

Therefore, I hypothesize:

Hypothesis 1a: The marginal difference in employee performance will be higher when both the supervisor and employee seek behavioral health support compared to when the employee seeks support, but the supervisor does not. (H1a)

However, an alternative explanation is that supervisors who seek behavioral health care may project their own vulnerabilities onto subordinates, potentially resulting in harsher evaluations. Supervisors who have experienced the challenges associated with behavioral health care might adopt a stricter evaluative stance, consciously or unconsciously, to distance themselves

from perceived weaknesses or to conform to organizational norms of resilience and stoicism. This perspective aligns with theories of self-protection and cognitive dissonance, where individuals may overcompensate for traits they perceive as stigmatized or counter-normative in themselves by being more critical of them in others (Harmon-Jones and Mills, 2019).

Organizations with strong cultures that emphasize strength and reliability might further incentivize supervisors to downplay the importance of behavioral health in evaluations, especially if they perceive it as incongruent with the organization's expectations of resilience. Thus, rather than fostering empathy and psychological safety, supervisors' own behavioral health experiences could amplify stigma and heighten scrutiny, particularly if they internalize organizational norms that view behavioral health as a weakness.

Therefore, I also hypothesize:

Hypothesis 1b: The marginal difference in employee performance will be lower when both the supervisor and employee seek behavioral health support compared to when the employee seeks support but the supervisor does not. (H1b)

Additionally, I expect that the intensity of the supervisor's behavioral health care engagement will further moderate this relationship. Frequent engagement in behavioral health care may signal a sustained commitment to personal well-being and resilience, which can shape supervisors' perceptions of wellness behaviors in their subordinates. Supervisors who consistently seek behavioral health support may develop a deeper understanding of the challenges associated with maintaining mental health, fostering a more empathetic and accepting evaluative stance. This consistent modeling of well-being prioritization can create a psychological climate where employees feel safer disclosing and addressing their mental health needs, thus potentially enhancing

their perceived commitment and performance (Barnes and Wagner, 2023).

In contrast, supervisors with lower intensity in behavioral health engagement may perceive wellness behaviors as less integral to professional performance, particularly in environments where such practices are not the norm. This sporadic engagement may weaken the perceived alignment between supervisor and subordinate, limiting the positive effects of shared wellness practices. Therefore, frequent and sustained behavioral health care engagement by supervisors is likely to reinforce a supportive evaluative context, particularly for employees who also prioritize their mental health.

Accordingly:

Hypothesis 2a: The marginal difference in employee performance for employees who seek behavioral health support will be higher when their supervisors seek behavioral health care at high intensity compared to when their supervisors seek behavioral health at low intensity. (H2a)

Alternatively, higher supervisor engagement could heighten stigma. Supervisors frequently using behavioral health support might project vulnerabilities onto subordinates, triggering self-protection mechanisms that result in stricter evaluations (Kaaronen, 2018). Supervisors utilizing extensive behavioral health resources may, paradoxically, evaluate similar help-seeking behaviors in subordinates more severely in structured military environments, as they project their own anxieties about how their above-average treatment conflicts with institutional emphasis on resilience.

Also, accordingly:

Hypothesis 2b: The marginal difference in employee performance for employees who seek behavioral health support will be lower when their supervisors seek behavioral health care at

high intensity compared to when their supervisors seek behavioral health at low intensity. (H2b)

3.3 Methods

3.3.1 Research Setting and Empirical Context

This study examines a panel of active-duty U.S. Army officers from January 2010 to May 2024. The Army provides a unique setting for studying employee-supervisor dyadic relationships due to its hierarchical structure, high-stress environment, and competitive "up-or-out" promotion system based on zero-sum evaluations that are a reflection of individual and peer performance.

The dataset captures the complex organizational hierarchy, with individuals appearing both as employees with bosses and potentially as a supervisor with subordinate employees. This creates a multi-level relational network spanning thousands of work units and over one hundred job classifications. Throughout the study period, personnel experience changes in positions, locations, and ranks.

My research focuses on behavioral health (BH) seeking patterns in supervisor-employee dyads over time. Using a rolling three-year window, I track whether employees' supervisors (first and second-line) sought BH services, and analyze how these patterns manifest in associated employee performance appraisal outcomes. The matching of supervisors and employees in this study is determined by the Army's structured placement system, designed to consider both individual employee preferences and organizational priorities. Officers typically switch jobs every one to three years. During each job placement cycle, employees rank available positions based on their preferences (e.g., the geographic location of the job, type of work unit). Simultaneously, organizations rank employee candidates based on the employee's qualifications, previous expe-

riences, and a virtual or in-person interview. After employees rank available jobs and organizations rank potential employees, an algorithm matches them, creating a slate of projected positions for the employees. During this market-sorting process, there is a stronger weight given to employee preferences. However, final employee-to-job matches are contingent on organizational requirements, available vacancies, and broader personnel management constraints. As a result, supervisor-employee dyads are formed in a manner that is structured and semi-random. Importantly, these relationships can be considered exogenous regarding behavioral health seeking tendencies, as neither supervisors nor employees have access to each other's medical records before or when new working relationships form.

3.3.2 Primary Sample and Variables

3.3.2.1 Sample

Figure 3.1 provides an overview of the data construction process used in the study, tracing how personnel, performance, and health appointment records were integrated to form the final analytic sample. Data from three primary sources—the Defense Manpower Data Center, the Army Evaluation Entry System, and the Defense Health Agency—were merged using a de-identified personal identifier (PID). Inclusion criteria limited the sample to active-duty Army officers in their third year of service and had two Army officer supervisors identified in their annual evaluations.[1] Supervisors were required to be documented first- and second-line raters. Individuals were then categorized based on their behavioral health care utilization into three groups: Never Seekers (no mental health appointments), Only Behavioral Health Seekers (outpatient mental health

[1]Supervisors could also be Department of Defense Civilian (non-uniformed) employees or officers from other branches of the armed forces (e.g. Air Force). However, I did not have the requisite personal and medical appointment data to retain these supervisors as observations in the study.

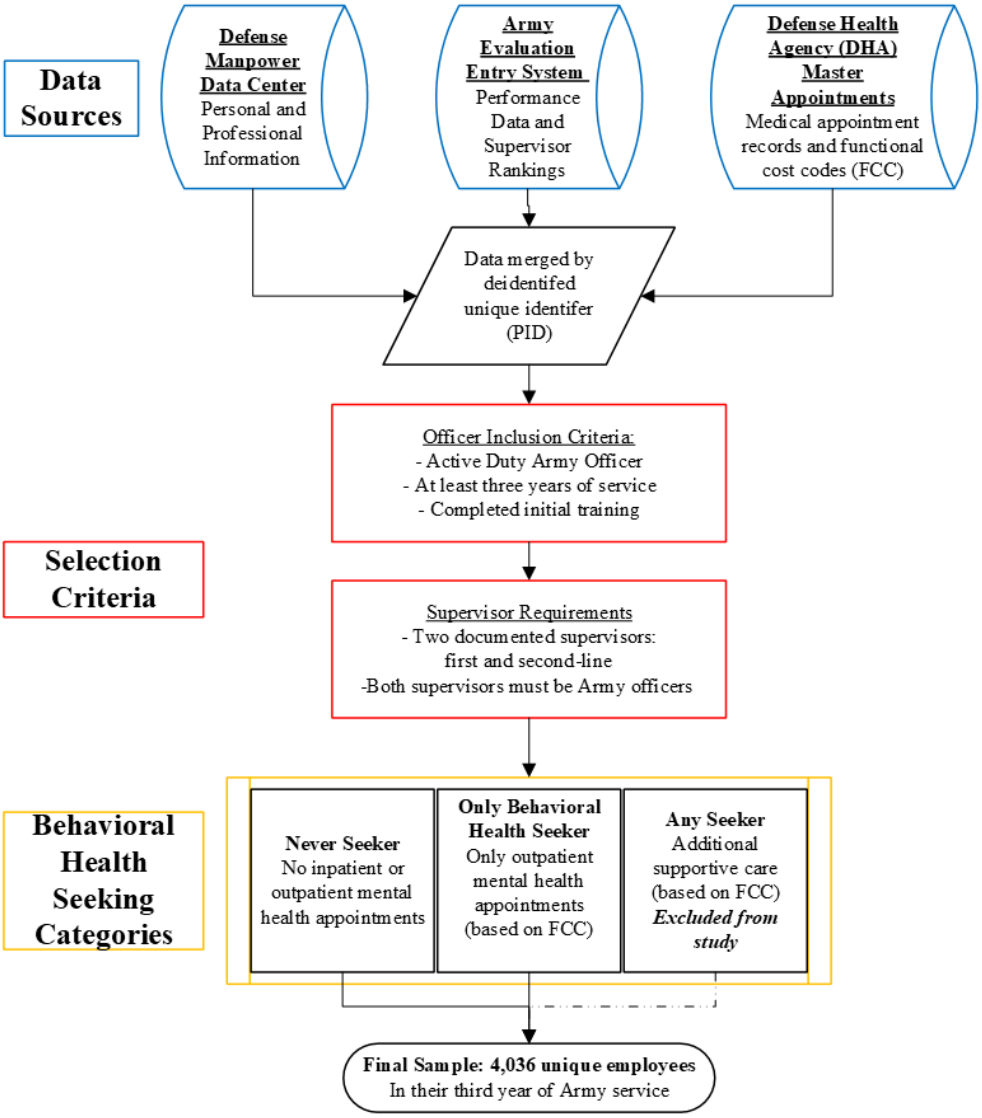

Figure 3.1: Data Sources and Sample Selection Process

appointments only), and Any Seekers (with broader care use, who were excluded from the analysis). This process yielded a final analytic sample of 4,456 officers in their third year of Army service. Further details on the sample selection process can be found in Appendix C.

3.3.2.2 Dependent Variables

The dependent variable, $Performance_{ijt}$, is a continuous standardized outcome of individual performance for the periods of observation. The data for this variable comes from the evaluation files and is a combined score of the first and second-line supervisors score for an individual. Typically, an individual only receives one performance report annually, unless other qualifying events occur that allow for additional reports (e.g., change of bosses, promotion opportunities, employee misconduct).

3.3.2.3 Key Explanatory Variable

The key explanatory variable, $BH.Seek_{jt}^{Dyad}$ is the BH-seeking behavior of the supervisors in the employee-supervisor dyad, as modeled in equation 3.1, I analyze the marginal difference of performance between employees that have a first and second line-supervisor who both sought BH in the last three years compared to employees whose supervisors do not. To accurately link supervisor behavioral health seeking behavior to employee performance evaluations, it is critical to establish a theoretically and empirically appropriate observation window. I use a three-year window to define supervisor and employee behavioral health seeking behavior to ensure that supervisors' behavioral health seeking occurred during the period when they were actively evaluating the employees included in the sample. Because all employees in the analytic sample have at least three years of active-duty service, this window captures relevant wellness behaviors

without extending into periods prior to their active military careers. Although no prior empirical studies have defined an optimal time window for linking supervisor behavioral health seeking to appraisal behaviors, this approach balances theoretical relevance and data availability. Further, in equation 3.2, I extend equation 3.1 by introducing a moderator for the supervisor's behavioral health-seeking intensity, $Intensity_{jt}$. Among individuals who sought behavioral health care in a given year, I calculated the mean number of appointments; individuals with appointment counts at or above the annual mean were classified as high-intensity seekers (coded as 1), and those below the mean were classified as low-intensity seekers (coded as 0).

3.3.2.4 Control Variables

To interpret the effects of the supervisor-employee dyads, it was necessary to control for other observed variables that may have impacted BH seeking behaviors. First, I include individual controls based on race, gender, and marital status. Second, I created a series of employee BH seeking controls, including if the employee ever seeks BH, if the employee had sought BH in the last three years, and if the employee ever had more than the mean number of BH appointments in a given year compared to other military officers. Also, I create a series of dummy variables to determine if the employee and 1st (and 2nd) line supervisor are of the same race, gender, or race and gender. Lastly, I consider the first- and second-line supervisor spans of control (e.g. how many subordinate employees the supervisor has).

3.3.3 Key Descriptive Statistics

The descriptive statistics table, Table 3.1, provides the mean, standard deviation, minimum and maximum for the sample. Additionally, Table 3.1 provides a mean and standard deviation for the

Table 3.1: Descriptive statistics for employees with 3 years of service

	Full Sample					Neither		Both		One	
Variables	Count	Mean	SD	Min	Max	Mean (Neither)	SD (Neither)	Mean (Both)	SD (Both)	Mean (One)	SD (One)
Composite Eval Score by Year Standardized (Z-Score)	4456	-0.039	1.025	-5.198	1.226	0.012	1.032	-0.081	0.907	-0.115	1.026
Received at least one 'most qualified' rating from 2LS	4456	0.596	0.491	0.000	1.000	0.610	0.488	0.586	0.494	0.574	0.495
All not white individuals (=1)	4456	0.279	0.449	0.000	1.000	0.275	0.447	0.279	0.449	0.285	0.452
All females (=1)	4456	0.202	0.402	0.000	1.000	0.184	0.388	0.263	0.441	0.223	0.416
Age (Years)	4456	26.917	2.916	22.000	55.000	26.832	2.771	26.956	3.075	27.048	3.110
Individuals not married (=1)	4456	0.611	0.488	0.000	1.000	0.622	0.485	0.562	0.497	0.601	0.490
Has only sought BH help; no other categories	4456	0.284	0.451	0.000	1.000	0.258	0.437	0.375	0.485	0.312	0.464
Individual had BH appt in last 3 years	4456	0.151	0.358	0.000	1.000	0.129	0.336	0.275	0.447	0.167	0.373
Individual had more than the mean BH appts (year)	4456	0.026	0.160	0.000	1.000	0.026	0.160	0.028	0.165	0.026	0.160
Rank Grouping	4456	1.002	0.050	1.000	2.000	1.002	0.048	1.004	0.063	1.002	0.050
Length of supervisor relationships (months)	4456	11.227	5.157	2.000	60.000	11.423	5.297	10.250	4.206	11.062	5.041
Employee and supervisor are same gender (binary)	4456	0.748	0.434	0.000	1.000	0.754	0.431	0.745	0.437	0.739	0.439
Employee and supervisor are same race (binary)	4456	0.633	0.482	0.000	1.000	0.631	0.483	0.641	0.481	0.634	0.482
Employee and supervisor match race and gender	4456	0.493	0.500	0.000	1.000	0.499	0.500	0.490	0.501	0.483	0.500
Employee and 2LS are same gender (binary)	4456	0.774	0.418	0.000	1.000	0.785	0.411	0.729	0.445	0.762	0.426
Employee and 2LS are same race (binary)	4456	0.618	0.486	0.000	1.000	0.629	0.483	0.637	0.482	0.597	0.491
Employee and 2LS match race and gender (binary)	4456	0.497	0.500	0.000	1.000	0.520	0.500	0.470	0.500	0.463	0.499
1st Line Supervisor Span of Control (count)	4456	4.387	3.307	1.000	49.000	4.231	3.019	4.825	4.032	4.572	3.600
2nd Line Supervisor Span of Control (count)	4456	23.751	14.095	1.000	101.000	24.880	14.001	21.876	17.525	22.220	13.466

Note: The table presents descriptive statistics for the full sample (N=4,456). The group means are identified by if their first or second line supervisor has sought behavioral health in the last three years. First and second line supervisor behavioral health seeking behavior is indicated by: neither supervisor has, one of the supervisors has, or both of them have. (Neither, Both, and One). SD = Standard Deviation. BH = Behavioral Health. 2LS = Second Line Supervisor.

three subgroups studied where neither supervisor has sought behavioral health in the last three years, both supervisors have sought behavioral health in the last three years, or either the first or second-line supervisor has sought behavioral health in the last three years. For performance scores, which are standardized for a comparative analysis, I see that the whole group, as well as the three subgroups, all have a mean close to zero and a standard deviation of one. Meaning that employees have a similar spread in performance scores regardless of supervisor behavioral seeking tendencies and at a descriptive level, this suggests I will not find an effect with the regressions.

The sample consists of 4,456 employees, with 2,881 unique first-line supervisors, and 1,053 unique second-line supervisors. The primary sample is approximately 28% not white, and has similar representation across the three subgroups. The sample is 21% female, and there is an increase in female representation across the three subgroups, with only 18% representation in the neither supervisor seeks group, and up to 26% representation in the both supervisors seek group. Within the sample, at three years of military service, 15% of employees in the sample had a behavioral health appointment within the last three years. Employee seeking behavior increases

with supervisor seeking behavior across subgroups, rising from 13% (neither seeking), to 17% (one seeking), to 28% (both seeking), in the previous three years.

3.3.4 Research Design

To test my theoretical claims, I analyze whether the behavioral health-seeking behavior of first- and second-line supervisors influences the annual performance scores of employees in their third year with the organization. As previously noted, the assignment of employees to their supervisors is exogenous to the supervisors' behavioral health engagement, given that neither employees nor supervisors have access to each other's medical records. This study focuses exclusively on supervisors who seek outpatient mental or behavioral health services, compared to those who do not utilize any behavioral wellness programs within the military healthcare system. Identification of supervisors was conducted using appointment billing reporting codes, which serve as the basis for sample selection. Additional details regarding the source and classification of these reporting codes are provided in Appendix C.

3.3.4.1 Model Specification

$$\begin{aligned}\text{Performance}_{ijt}|BH.Seek_j = \alpha_i + \beta_1 \left(BH.Seek_{jt}^{Dyad}\right) \\ + \beta_2 Controls_{ijt} + \gamma_t + \lambda_t + \varepsilon_{ijt}\end{aligned} \tag{3.1}$$

$$\begin{aligned}\text{Performance}_{ijt}|(BHSeek_j = 1) = \alpha_i + \beta_1 \left(BHSeek_{jt}^{Dyad}\right) \\ + \beta_2 Intensity_{jt} \\ + \beta_3 Controls_{ijt} + \gamma_t + \lambda_t + \varepsilon_{ijt}\end{aligned} \tag{3.2}$$

Where $Performance_{ijt}$ is the performance score for employee i at time t, with supervisor j. The key explanatory variable, $BHSeek_{jt}$, is the BH seeking behavior of the supervisors within the employee-supervisor dyad. within the last t years. The model includes a set of control variables, $Controls_{ijt}$, that account for individual and employee-supervisor dyad characteristics, and occupational workgroups. The model also includes year fixed effects to control for temporal trends. Finally, ε_{ijt} is the error term. In equation 3.2, among all supervisors who ever seek BH, I include the intensity of the supervisor's BH seeking behavior, $Intensity_{jt}$, as a moderator to the model. The model is estimated using ordinary least squares (OLS) regression, with robust standard errors clustered at the employee level.

3.4 Results

3.4.1 Primary Results

Table 3.2 reports the results. Overall, I find results are inconclusive to support Hypothesis 1a, and I find limited support for Hypothesis 1b. Specifically, for hypothesis 1b, when one of an employee's supervisors seek behavioral health care compared to when no supervisors seek care, the employee experiences a decrease in performance evaluation. For hypothesis 2a and 2b, overall the results are inconclusive.

The primary model (Table 3.2, Model (1)) examines the extensive margin of supervisor behavior, focusing on a strict comparison between dyads with full supervisor behavioral health engagement (both supervisors) and dyads with none, among employees who received behavioral health in the last three years. Employees with only one supervisor (either first or second-line) seeking behavioral health were omitted from this analysis. In contrast, the second model(Table 3.2,

Model (2)) explores a robust intensive margin that captures partial exposure by comparing dyads where at least one supervisor sought behavioral health care to those with no exposure. The third model (Table 3.2, Model (3)) narrows the intensive margin analysis by excluding employees with non-seeking supervisors, instead comparing performance between employees whose supervisors both sought behavioral health services against those with only one supervisor doing so. The fourth model (Table 3.2, Model (4)) moderates model (1), by focusing on first-line supervisors with high-intensity appointment utilization that year. Specifically, high-intensity behavioral health seeking was defined relative to the annual distribution of appointment frequency. The mean was selected as the cutoff point to reflect typical behavioral health engagement levels among seekers in each year, balancing the need to capture meaningful differences in appointment intensity while maintaining sufficient sample size for high-intensity classifications. The fifth model (Table 3.2, Model (5)) moderates model (2), by dropping never-seeking supervisors and focusing on the difference in first line supervisor behavioral health seeking intensity for supervisor-employee dyads where at least one supervisor has sought behavioral health in the last three years.

At the extensive margin, Table 3.2 Model (1), shows the differential effect of a first and second-line supervisor seeking behavioral health on an employee's performance is inconclusive when the employee has three years of employment in the organization. In Model (2) of Table 3.2, comparing cases where either the first- or second-line supervisor sought behavioral health care again yields inconclusive results between dyads where both versus neither supervisor sought care. However, in Model (2) (using an OLS regression) if one of the supervisors has sought behavioral health, I note a .09 standard deviation decrease in employee performance compared to supervisors who do not seek BH ($\beta = -0.092, SE = 0.032$). In Model (3) of Table 3.2, I observe the

Table 3.2: Primary regression results for employees with 3 years of service

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 3	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 3	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 3	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 3	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 3
IV: Extensive Margin Indicator (BH in 3 years) = 1	0.000			-0.218	
	(0.061)			(0.161)	
IV: Robust Intensive Margin Indicator (BH in 3 years) = 1		-0.021			
		(0.060)			
IV: Robust Intensive Margin Indicator (BH in 3 years) = 2		-0.092***			-0.056
		(0.032)			(0.063)
IV: Narrow Intensive Margin Indicator (BH in 3 years) = 1			0.050		
			(0.062)		
1LS had more than mean BH appointments = 1					-0.096
					(0.090)
Individual had BH appt in last 3 years = 1	-0.417***	-0.326***	-0.214***	-0.437***	-0.215***
	(0.067)	(0.052)	(0.073)	(0.074)	(0.072)
Indvidual had more than the mean BH appts (year) = 1	-0.217	-0.238**	-0.174	-0.249*	-0.174
	(0.135)	(0.109)	(0.163)	(0.142)	(0.163)
All not white individuals (=1) = 1	-0.180***	-0.218***	-0.283***	-0.178***	-0.280***
	(0.047)	(0.037)	(0.056)	(0.049)	(0.056)
All females (=1) = 1	0.069	0.088*	0.085	0.078	0.084
	(0.061)	(0.047)	(0.066)	(0.066)	(0.066)
Individuals not married (=1) = 1	-0.020	-0.062**	-0.126***	-0.017	-0.125***
	(0.039)	(0.031)	(0.047)	(0.041)	(0.047)
Rank Grouping = 2, 2	-0.217	-0.198	-0.334	-0.071	-0.339
	(0.370)	(0.319)	(0.527)	(0.397)	(0.528)
Employee and supervisor are same gender (binary) = 1	-0.071	-0.031	0.030	-0.071	0.030
	(0.072)	(0.056)	(0.080)	(0.077)	(0.080)
Employee and supervisor are same race (binary) = 1	-0.014	-0.026	-0.014	-0.022	-0.011
	(0.077)	(0.059)	(0.083)	(0.083)	(0.083)
Employee and supervisor match race and gender = 1	0.071	0.043	-0.053	0.089	-0.055
	(0.089)	(0.069)	(0.098)	(0.095)	(0.098)
Span of control (1LS), Standardized	-0.022	-0.044***	-0.065***	-0.024	-0.066***
	(0.015)	(0.011)	(0.015)	(0.016)	(0.015)
Span of control (2LS), Standardized	0.026*	0.029**	0.029	0.032**	0.028
	(0.015)	(0.012)	(0.018)	(0.016)	(0.018)
Occupational workgroup = 2, Operations Support	-0.184	-0.138	-0.171	-0.175	-0.172
	(0.174)	(0.134)	(0.183)	(0.193)	(0.183)
Occupational workgroup = 3, Force Sustainment	0.063	0.090	0.074	0.083	0.075
	(0.080)	(0.062)	(0.091)	(0.085)	(0.091)
Occupational workgroup = 4, Medical	0.067	0.039	-0.003	0.057	-0.001
	(0.091)	(0.069)	(0.099)	(0.096)	(0.099)
Occupational workgroup = 5, Other	0.165**	0.132**	0.075	0.164**	0.076
	(0.070)	(0.054)	(0.080)	(0.073)	(0.080)
Total supervisor relationship length, Standardized	0.135***	0.130***	0.119***	0.128***	0.120***
	(0.028)	(0.023)	(0.038)	(0.029)	(0.038)
Presidential Administration = 2, Trump	-0.095*	-0.098**	-0.128**	-0.078	-0.125**
	(0.051)	(0.040)	(0.059)	(0.054)	(0.059)
Presidential Administration = 3, Biden	-0.061	-0.039	-0.036	-0.050	-0.037
	(0.053)	(0.042)	(0.061)	(0.056)	(0.061)
Constant	-0.017	0.041	0.077	-0.040	0.136
	(0.109)	(0.086)	(0.121)	(0.117)	(0.133)
Observations	2,848	4,456	1,859	2,625	1,859
R-squared	0.047	0.047	0.049	0.048	0.050
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents primary regression results for employees with 3 years of service. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

narrow intensive margin by dropping observations for when both supervisors do not seek BH. The impact on employee performance when comparing employees who have one supervisor who seeks BH from an employee that has two are inconclusive.

Analyzing the results of Models (4) and (5) of Table 3.2 show that supervisor's intensity of BH appointments, both at the extensive margin and intensive margin, have an inconclusive effect on employee performance. However, the results across all five models show that seek-

ing behavioral health is detrimental to your performance evaluations at a statistically significant rate. The smallest observed effect on employee performance is observed in Model (3) ($\beta = -0.214, SE = 0.073$) and (5) ($\beta = -0.215, SE = 0.072$), respectively. The largest negative effect is seen at the extensive margins of Model (1) ($\beta = -0.417, SE = 0.067$) and (4) ($\beta = -0.437, SE = 0.074$), respectively.

Additionally, across all five models, employees that are not white receive statistically significant lower performance scores on average than their white peers ($\beta = -0.178$ to $-0.283, SE = 0.049$ to 0.056). Only one control variable yielded a significant positive effect on performance. That is, having a supervisor for a one standard deviation longer period than average had an average marginal effect of a $\beta = 0.119$ (Model (3),$SE = 0.038$) to $\beta = 0.135$ (Model (1), $SE = 0.028$) standard deviation increase in performance score. There is also a negative average marginal effect on employee performance scores across models (2), (3), and (5) for not being married, and in models (2) and (4) as first line supervisor span of control increases.

3.4.2 Heterogeneity and Temporal Extensions

As a heterogeneity check, I rerun the primary model with employees that have seven and eleven years of service respectively. The descriptive statistics and results for these models can be found in Appendix D. I find similar trends to the primary model, with an exception that seeking behavioral health as an employee has a slightly decreased, although still negative, effect on performance ($\beta = -0.204, SE = 0.099$ to $\beta = -0.210, SE = 0.076$) than the primary model in Table D.2. Also, in these results the performance penalty for being non-white is no longer significant. Interestingly, in model (4) at seven years of service, I find that the first line supervisor's intensity of BH appointments has a positive effect on employee performance ($\beta = 0.741, SE =$

0.350). This is the only model where I observe a positive effect of supervisor BH seeking on employee performance. At eleven years of service, I find a divergence to the primary model. At the extensive margins, Table D.4, Model (1) and Model (4) show an increasingly negative effect on the marginal effect of employee BH seeking on performance scores, with a $\beta = -0.536, SE = 0.183$ and $\beta = -0.609, SE = 0.194$, respectively. There is still a moderate negative effect when evaluating at the robust intensive margin, Table D.4, Model (2), with a $\beta = -0.319, SE = 0.147$. However, the intensive margin of Model (3) and Model (5) show an inconclusive effect on employee performance scores. This suggests that at eleven years of service, employees are penalized for seeking BH services when their supervisors do not seek BH services, and the results are inconclusive for seeking BH services when their supervisors do.

3.4.3 Robustness Check

I further stratified the model to examine the effects of supervisor behavioral health seeking behavior among only employees who have received behavioral health services in the last three years. This analysis provides similar results to the primary model showing the full sample. Similarly, when I stratify the data to include only employees who have not sought behavioral health in the last three years, and never seek behavioral health support in the dataset, I find similar results to my main models.

3.4.4 Sensitivity Checks

To check the model specifications, I rerun the primary models with a binary indicator for employee performance scores where employees, during any year at their current rank, who receive at least one 'most qualified' rating from their second line supervisor are coded as 1, and all oth-

ers are coded as 0. I run this sensitivity check for employees with three, seven, and 11 years of service. Due to concerns about statistical conclusion validity, I do not include the results for employees with 11 years of service, as the sample size is too small to draw reliable conclusions. The results for employees with three and seven years of service are shown in Table D.11 and Table D.12, respectively.

3.4.4.1 Sensitivity Check for Three Years of Service

The results show that the model is robust to this specification, with similar magnitude and direction to the variables in the primary model, at three years of service. Additionally, Table D.11 shows that there is an additional negative average marginal effect for employees that have more than the mean BH appointments per year, with the greatest effect occurring on the extensive margins of Model (1) ($\beta = -0.203, SE = 0.057$) and (4) ($\beta = -0.215, SE = 0.073$), respectively. At the intensive margins, the results are still negative, but statistically lower than the extensive margins. An additional nuance of this operationalization can be seen in, Table D.11, Model (5), where the employee's chance of getting a most qualified rating from there second line supervisor decreases by almost 12% when their first line supervisor has more than the mean number of BH appointments per year ($\beta = -0.118, SE = 0.057$). This suggests that the first line supervisor's BH seeking behavior has a negative effect on the employee's performance rating, and that the first line supervisor's BH seeking behavior is more important than the second line supervisor's.

3.5 Discussion

These findings raise important questions about the disconnect between organizational rhetoric and actual practices regarding employee well-being. These findings align with previous research

on stigma in organizational settings (Corrigan, 2000; Campbell et al., 2023). Despite institutional efforts to normalize behavioral health support, seeking help specifically, or absenteeism more broadly, appears to create what Kreiner et al. (2022) describe as boundary conditions separating employees into out-groups. This suggests that attempts to create psychologically safe environments (Edmonson, 1999) may be undermined by implicit supervisor bias in performance evaluation systems.

The temporal analysis reveals that supervisors' perception of employees seeking wellness care, or more generally absenteeism, persist and potentially worsen as the employee's longevity increases in the company. The performance penalty increases from three years of service ($\beta = -0.417, SE = 0.067$) to eleven years of service ($\beta = -0.536, SE = 0.183$), suggesting that seeking help (or just being absent) becomes more professionally costly. These findings suggest that behavioral health stigma may not simply persist over time but compound, entrenching disadvantages across critical career junctures. This aligns with Barnes et al. (2023) observation that organizational systems often create tensions between performance demands and human sustainability needs.

The study's finding that employee performance improves with longer supervisor relationships ($\beta = 0.119$ to 0.135) suggests that relationship stability may help mitigate some negative effects. However, this benefit appears insufficient to overcome the substantial penalties associated with seeking behavioral health support. This tension exemplifies what Barnes and Wagner (2023) describe as the challenge supervisors face of balancing organizational performance with employee well-being.

Particularly troubling is the compounding effect on historically marginalized groups. The results show that non-white employees face additional performance penalties early in their ca-

reers, independent of behavioral health seeking behavior ($\beta = -0.178$ to $-0.280, SE = 0.037$ to 0.056). This intersectionality of race and wellness stigma supports previous findings about how seemingly neutral evaluation systems can perpetuate systemic inequities (Castilla and Benard, 2010). The disproportionate penalties experienced by racial minority employees in performance evaluations suggest that behavioral health stigma may not operate uniformly across employees but instead interact with other marginalized identities. These findings are consistent with broader theories of intersectional disadvantage, which argue that marginalization based on race, health, and other social identities often interact to produce compounding disparities in organizational outcomes (Evans et al., 2024). Recent methodological work emphasizes the importance of accounting for these layered inequalities, as traditional approaches may underestimate the combined effects of intersecting identities on evaluative judgments.

3.5.1 Mechanisms

One explanation for these findings is rooted in stigma persistence and the creation of boundary conditions that separate employees into perceived in-groups and out-groups. Even when supervisors seek BH support themselves, the infrequency of such engagements (e.g., once or twice a year) may prevent meaningful attitudinal shifts that influence how they perceive mental health needs in others. Without regular engagement, the experience of seeking support is less likely to challenge deep-seated stigmas, leading supervisors to maintain stereotypical views of wellness-seeking employees as less resilient or committed (Corrigan, 2000; Campbell et al., 2023). Furthermore, supervisors often do not know if their subordinates are seeking BH care, which means their evaluations are influenced by visible behaviors, such as absenteeism or altered schedules, that may be misattributed to a lack of motivation rather than legitimate wellness needs. This

aligns with Kreiner's concept of stigmatized work, where behaviors perceived as deviations from organizational norms of strength and reliability mark employees as part of an "out-group," subtly influencing performance assessments (Kreiner et al., 2022).

Another potential mechanism is norm enforcement within hierarchical structures. In military and paramilitary organizations, cultural expectations around resilience and self-sufficiency are deeply ingrained. Supervisors who seek BH support once or twice a year may compartmentalize their own wellness activities as private and separate from their role as evaluators. This infrequent engagement may allow them to conform publicly to organizational norms of toughness and reliability, reinforcing these expectations during evaluations. Unaware of their subordinates' BH status, supervisors may interpret perceived disengagement or inconsistencies like missed deadlines or altered schedules as a lack of commitment rather than as adjustments related to health. This phenomenon is consistent with role congruity theory, which posits that supervisors may be stricter with employees who deviate from established norms of reliability and resilience, even if those deviations are linked to legitimate wellness needs (Eagly and Karau, 2002).

An additional avenue for exploration involves disaggregating the sample based on whether employees themselves seek behavioral health care. Doing so may help clarify whether supervisors respond differently to wellness-seeking behaviors depending on if the behavior is known, inferred, or assumed. While this stratification could provide deeper insight into how relational dynamics and performance evaluations interact, it also introduces potential selection bias, as employees who seek BH support may systematically differ from those who do not. Both in observable characteristics and unobserved traits such as coping style, perceived stigma, or prior performance. Nevertheless, examining this subsample may reveal whether the presence of a similar need to seek care between supervisor and subordinate moderates evaluative outcomes in more

nuanced ways.

3.5.2 Theoretical Implications

This study advances several key theoretical contributions. First, it extends Restricted Employee Sustainability Theory (REST) by demonstrating that supervisor behavioral health seeking patterns alone do not create the supportive environments theorized to enhance employee sustainability. The persistent negative performance effects for employees seeking help, regardless of supervisor behavior, suggests that organizational structures and evaluation systems may override individual leader influences.

Second, these findings challenge existing theoretical assumptions about psychological safety and leader modeling behavior. While prior work suggests that leader vulnerability around wellness needs should normalize help-seeking (Edmonson, 1999), my results indicate this may not translate to formal performance evaluations. This expands our understanding of the limitations of psychological safety in hierarchical settings.

Third, the results contribute to merit-based evaluation theory by revealing how wellness-seeking behaviors become inadvertently penalized even in systems designed to be objective. This builds on Castilla and Benard (2010) work by showing how performance evaluation systems can structurally disadvantage employees who prioritize sustainability practices officially endorsed by the organization.

Finally, the temporal analysis showing increasing penalties over career progression advances theoretical understanding of how sustainability stigmas might compound over time. This suggests that wellness-seeking behavior, specifically, or absenteeism, more broadly, may not be viewed equally based on employees' career stage and organizational responsibilities (Kreiner

et al., 2022).

3.5.3 Practical Implications

This research has several important practical implications for organizations and leaders. First, despite explicit organizational messaging promoting behavioral health support, the significant performance penalties for employees who seek help reveals a concerning misalignment between rhetoric and reality. Organizations must recognize that simply encouraging wellness seeking behavior is insufficient. They must actively reform evaluation systems that potentially penalize such choices.

Second, the finding that supervisor behavioral health seeking patterns do not mitigate negative effects on employee evaluations suggests that cultural change requires more than leader modeling. Organizations need to implement structural changes to performance management systems, potentially including explicit criteria for how wellness-seeking behaviors should be considered in evaluations (Castilla and Ranganathan, 2020). This is somewhat contrary to the findings of Dimoff and Kelloway (2019) who found that supervisors showed more consideration and support for employees seeking wellness support.

Third, the compounding negative effects over career progression indicate that organizations may be inadvertently pushing out employees who prioritize sustainability. This suggests a need for longitudinal tracking of career outcomes related to behavioral health seeking to identify and address systematic barriers to advancement.

Fourth, the additional performance penalties faced by non-white employees highlight how behavioral health stigma can intersect with and amplify existing workplace inequities. Organizations must take an intersectional approach to wellness support that considers how different em-

ployee groups may face compounded barriers.

Finally, while longer supervisor relationships show modest positive effects on performance, these benefits are overshadowed by behavioral health seeking penalties. This suggests organizations should focus on creating systemic support for wellness rather than relying primarily on relationship-based solutions.

3.5.4 Limitations and Future Research Directions

While this study provides important evidence of behavioral health stigma within supervisor-employee evaluations, it has several limitations that future research could address. First, although the observed patterns hint at intersectional effects, the current study does not formally model heterogeneity across race, gender, and behavioral health status simultaneously. Future research could apply intersectional multilevel modeling frameworks, such as multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA), to better capture how layered identities shape performance outcomes (Evans et al., 2024).

Second, while this analysis focuses on behavioral health care seeking as a proxy for well-being, future work could extend to other wellness-related behaviors (e.g., physical health maintenance, counseling participation) to test the generalizability of the findings. One way to evaluate this would be to look at self-reports during wellness checks compared to individual annual self-reports to primary care providers. By diving into the Behavioral Health Data Portal (BHDP) data, I could analyze at a greater scale the reason why people are most often seeking behavioral health support, and how they are rating their own wellness and resilience at the time of their appointment. Additionally, I could compare this data against their self-reported wellness data reported in their Periodic Health Assessment (PHA) which is evaluated by their primary care provider. This

would allow me to see how consistently they are reporting their wellness and resilience across multiple data points, and how this compares to their performance evaluations, as well as how they are rating their subordinates.

Third, a significant limitation of this study is that it does not capture the full range of behavioral health seeking options available to military personnel. In addition to seeking care at military treatment facilities, employees can seek care through reference to civilian providers, chaplains, or through third-party counselors like the Military Family Life Consultants (MFLC) program or Military OneSource. While the latter sources are inaccessible from a data perspective, I could leverage insurance reimbursement data for referred care, to non-military providers, to see if there are any differences in the performance evaluations of employees who seek care through these channels. Or, at a minimum, make my wellness seeking data more robust from an observation count. Additionally, the civilian data would allow me to see if any family members are also seeking wellness care in addition to the Solider, to observe if there are any spillover effects into employee or subordinate performance evaluations.

Finally, longitudinal studies tracking career trajectories and wellness outcomes could illuminate the long-term implications of early career behavioral health seeking decisions. Understanding these patterns could help organizations better support sustainable career development while maintaining effectiveness.

3.6 Conclusion

This study provides compelling evidence that employees face substantial performance penalties for seeking behavioral health support, irrespective of their supervisors' own behavioral health engagement. These findings highlight a critical disconnect between organizational wellness mes-

saging and actual evaluation outcomes. Despite claims of support for employee well-being, supervisor engagement with behavioral health services, whether occasional or more frequent, does not appear to mitigate the negative performance impacts experienced by employees who seek help. This gap suggests that supervisors, even when accessing similar wellness resources, may still perpetuate biases that penalize behavioral health engagement among their subordinates.

The results challenge prevailing assumptions about the role of supervisors in modeling wellness behaviors. While organizational rhetoric often emphasizes leader-driven normalization of wellness support, the evidence shows that mere supervisor participation in behavioral health services is insufficient to shield employees from stigma and its career consequences. This underscores the need for organizations to move beyond leader modeling and address the deeper structural biases embedded within performance evaluation systems. Simply put, supervisor behavior alone is not enough to disrupt the stigma associated with seeking support; systematic change is necessary to realign evaluation practices with organizational wellness goals.

These findings have important implications for both theory and practice. Theoretically, they extend our understanding of how supervisory influence interacts with formal evaluation systems to undermine psychological safety and organizational sustainability initiatives. Practically, they reveal the need for organizations to reform performance management approaches that inadvertently penalize wellness-seeking behaviors, even when supervisors themselves engage in those same practices. Furthermore, the compounded penalties experienced by non-white employees highlight the intersection of wellness stigma with broader workplace inequities, suggesting that supervisory biases may exacerbate disparities in career advancement.

Moving forward, organizations must work to align their stated values around employee well-being with their evaluation practices. This may require fundamentally rethinking how per-

formance is measured and rewarded to create environments where seeking support for behavioral health genuinely supports rather than hinders career progression. Only through such systematic change can organizations build truly sustainable workplaces that balance high performance with employee well-being.

Chapter 4

CONCLUSION

This dissertation offers critical insights into how organizational policies and supervisor behaviors interact to shape equity and advancement outcomes in hierarchical settings. Through two empirical studies, one on partial blindness in promotion evaluations and the other on the impact of supervisors' behavioral health-seeking behavior, this work underscores the central but often underappreciated role of supervisors in mediating policy effectiveness and perpetuating or mitigating structural inequality (Tsui et al., 1992; Kreiner et al., 2022).

The first study reveals the limited impact of partial blindness policies on promoting fairness in internal promotion decisions. While these policies aim to mask race and gender identifiers, the persistence of demographic cues, such as names and pronouns, allows evaluators to continue inferring candidates' identities (Norton et al., 2004; Eliezer et al., 2024). As a result, the policy fails to meaningfully alter promotion outcomes, particularly for underrepresented groups (Castilla, 2008). The data show that white and non-white men benefited more from the policy than women, especially non-white women. These findings challenge the assumption that removing surface-level demographic markers is sufficient to ensure meritocratic outcomes (Rosenbaum, 1979). Instead, they reveal the entrenched nature of relational demography (Tsui and O'Reilly III,

1989) and the compounding impact of early career advantages and evaluator similarity on long-term career trajectories (Forbes, 1987).

The second study focuses on supervisors' behavioral health (BH) usage and its association with performance evaluations. Contrary to theoretical expectations drawn from attribution theory and human sustainability leadership frameworks (Harmon-Jones and Mills, 2019; Barnes and Wagner, 2023), supervisors who seek BH support do not provide more empathetic or generous performance ratings. Nonetheless, the analysis uncovers a persistent performance disadvantage for racial minorities, suggesting that systemic biases override any potential mitigating effects of supervisors' self-care behaviors (Corrigan, 2000; Campbell et al., 2023)

Taken together, these studies demonstrate that both structural interventions (e.g., blindness policies) and interpersonal dynamics (e.g., supervisor behaviors) have limited standalone impact on achieving equity in supposedly meritocratic advancement systems (Castilla and Ranganathan, 2020). Instead, this dissertation argues for a more integrated approach that aligns institutional policy design with relational and contextual realities (Edmonson, 1999). This dissertation contributes to several streams of literature, including organizational behavior, leadership, and performance management. It extends existing research by examining the nuanced limitations of policies that are often presumed effective and by empirically testing how identity and supervisor-subordinate dynamics influence outcomes in high-stakes, hierarchical environments (Tsui et al., 1992; Eagly and Karau, 2002) . In doing so, it offers actionable insights for scholars and practitioners seeking to build more meritocratic and sustainable organizations.

Ultimately, this work underscores a crucial theme: formal policies alone cannot overcome informal structures and cultural inertia (Kreiner et al., 2022; Castilla, 2008). Achieving genuine meritocracy in promotion and evaluation systems requires coordinated efforts at both the pol-

icy and practice levels, recognizing that fairness is not merely a matter of removing bias at the surface but addressing the deeper systems through which bias is reproduced (Rosenbaum, 1979; Tsui and O'Reilly III, 1989).

Appendix A

PRIMARY RESULTS AND ROBUSTNESS CHECKS FOR CHAPTER 2

List of Variables, Functional Form, and Data Source

Table A.1 provides a list of variables, their functional form, and the data source for the analysis in Chapter 1. The table includes the dependent variable, key explanatory variables, and control variables used in the analysis. The data sources include the Defense Manpower Data Center Master File (DMDC Master File) and the Evaluation Entry System or Officer Evaluation Record System.

Table A.1: List of Variables, Functional Form, and Data Source

Variable	**Functional Form**	**Source; Detail**
Dependent Variable		
Promotion to lieutenant colonel	Binary	Defense Manpower Data Center Master File (DMDC Master File); selection for promotion from major to lieutenant colonel for full sample, males, females, whites, and non-whites
Key Explanatory Variables		
Implementation of the Partial Blindness Policy	Binary	Imputed (based on promotion timeline after policy implementation); Pre-period defined as officers considered for promotion to lieutenant colonel 2018-2020, post-policy for officers considered for promotion in 2021-2023
Controls		
Not White	Binary	DMDC Master File; Binary indicator of race (1 indicates not white; 0 indicates white)
Not Female	Binary	DMDC Master File; Binary indicator of gender (1 indicates not-female; 0 indicates male)
Age When Promoted to Major	Continuous	DMDC Master File; Years of Age when promoted to Major
Individuals Married	Binary	DMDC Master File; Individual Married when eligible for promotion to lieutenant colonel (1 indicates married; 0 indicates never married, divorced, or widowed)
Years of Active Federal Military Service (AFMS)	Continuous	DMDC Master File; Years of Active Federal Military Service when eligible for promotion to lieutenant colonel
Officer has completed ILE or SSC	Binary	DMDC Master File; Officer has completed expected professional military education for promotion to lieutenant colonel (e.g., Intermediate Level Education (ILE) and/or Senior Service College (SSC) (1 indicates completion of ILE and/or SSC; 0 indicates neither complete)
Any Joint Professional Military Education	Binary	DMDC Master File; Officer has completed any level of joint professional military education during their career (1 indicates yes; 0 indicates no)
Command Experience (Any Rank)	Binary	Primary: DMDC Master File, Secondary: Evaluation Entry System; Officer has commanded a military organization at any rank prior to being promoted to lieutenant colonel (1 indicates in command or previously commanded; 0 indicates never commanded)
Average Senior Rater (SR) Score (Standardized)	Continuous	Evaluation Entry System or Officer Evaluation Record System; Officers average senior rater score while at the rank of major; mean of 0, standard deviation of 1
Average score or Rater and SR (Standardized)	Continuous	Evaluation Entry System or Officer Evaluation Record System; Officers average composite rater and senior rater score while at the rank of major; mean of 0, standard deviation of 1

Logit Regression Results

This section presents the logistic regression results for four different samples, based on how the promotion board cohorts employees. Table A.2 shows the results for the core operations employees. The second table, Table A.3, presents results for intelligence personnel and communications staff. The third table, Table A.4 focuses on administrative personnel, human resources staff, and logistics employees. The last table, Table A.5 provide results for medical administration employees and nurses. These results supplement the main estimates for all variables presented in Tables 2.2 and 2.3.

Table A.2: Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, by race and gender (Core Operations Employees)

	(1)	(2)	(3)	(4)
	Logit ITS	Logit ITS	Logit ITS	Logit ITS
	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men Operations	Non-White Men Operations	White Women Operations	Non-White Women Operations
Post Policy: post == 1 = 1	0.056***	0.087***	0.060	0.095
	(0.013)	(0.031)	(0.041)	(0.096)
Age when promoted to Major	-0.000	-0.001	-0.002	-0.015**
	(0.001)	(0.003)	(0.013)	(0.008)
Individuals married = 1	0.017	0.017	0.030**	0.110
	(0.015)	(0.026)	(0.015)	(0.070)
Average Race Similarity to 1LS and 2LS	-0.028	0.009	-0.024	0.187
	(0.032)	(0.079)	(0.045)	(0.346)
Average Gender Similarity to 1LS and 2LS	-0.026	0.017	-0.049	-0.843
	(0.043)	(0.036)	(0.095)	(0.553)
Years of Active Federal Military Service	0.004**	0.007**	0.009	0.034**
	(0.002)	(0.003)	(0.016)	(0.017)
Officer has completed ILE or SSC = 1	0.104***	0.093***		0.243**
	(0.022)	(0.019)		(0.115)
Any Joint Professional Military Education = 1	0.008	-0.029	-0.077***	
	(0.013)	(0.036)	(0.028)	
Years from first rating officers to eligibility for LTC	-0.014*	-0.001		
	(0.007)	(0.012)		
Years as a supervisor of officers	0.035***	0.103***		
	(0.010)	(0.026)		
Total number of evaluations as supervisor (logged)	-0.001	-0.058*		
	(0.017)	(0.031)		
Branch similarity to supervisors	-0.075***	-0.153***	-0.026	-0.030
	(0.025)	(0.047)	(0.046)	(0.267)
Average Senior Rater Score (Scale:1-4)	0.203***	0.259**	0.188***	-0.434**
	(0.031)	(0.116)	(0.064)	(0.176)
Average Score of Rater and SR (Scale: 2-8)	0.152***	0.216***	0.122***	0.560***
	(0.018)	(0.061)	(0.046)	(0.118)
Year Promoted to Major	-0.023***	-0.029***	-0.009	-0.014
	(0.004)	(0.008)	(0.013)	(0.014)
Observations	3,572	917	151	76
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, by race and gender (Intelligence and Communications Employees)

	(1)	(2)	(3)	(4)
	Logit ITS	Logit ITS	Logit ITS	Logit ITS
	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men Ops Spt	Non-White Men Ops Spt	White Women Ops Spt	Non-White Women Ops Spt
Post Policy: post == 1 = 1	0.002	-0.026	0.045	-0.063**
	(0.016)	(0.050)	(0.085)	(0.031)
Age when promoted to Major	-0.002	0.000	-0.006*	0.032***
	(0.002)	(0.006)	(0.003)	(0.007)
Individuals married = 1	-0.003	0.011	0.037	0.065**
	(0.017)	(0.044)	(0.051)	(0.026)
Average Race Similarity to 1LS and 2LS	-0.008	0.025	0.090	0.109**
	(0.082)	(0.064)	(0.077)	(0.042)
Average Gender Similarity to 1LS and 2LS	-0.106	-0.158***	-0.072	-0.774***
	(0.082)	(0.058)	(0.126)	(0.245)
Years of Active Federal Military Service	0.002	0.004	0.007	-0.023***
	(0.003)	(0.007)	(0.008)	(0.007)
Officer has completed ILE or SSC = 1	0.120***	0.061*	-0.106***	0.235***
	(0.035)	(0.032)	(0.041)	(0.045)
Any Joint Professional Military Education = 1	0.002	0.044		-0.098***
	(0.024)	(0.050)		(0.037)
Years from first rating officers to eligibility for LTC	0.014*	0.001		
	(0.008)	(0.005)		
Years as a supervisor of officers	0.008	-0.013		
	(0.013)	(0.035)		
Total number of evaluations as supervisor (logged)	-0.024	0.055		
	(0.025)	(0.058)		
Branch similarity to supervisors	0.008	-0.006	0.079	0.106**
	(0.029)	(0.043)	(0.115)	(0.043)
Average Senior Rater Score (Scale:1-4)	0.220***	0.431**	-0.236	0.377***
	(0.072)	(0.171)	(0.191)	(0.122)
Average Score of Rater and SR (Scale: 2-8)	0.169***	0.123	0.345***	0.346***
	(0.042)	(0.097)	(0.087)	(0.034)
Year Promoted to Major	-0.014*	-0.001	-0.003	0.009
	(0.007)	(0.014)	(0.015)	(0.012)
Observations	991	419	131	138
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, by race and gender(Admin, HR, and Logistics Employees)

VARIABLES	(1) Logit ITS dv_promo White Men Force Sust	(2) Logit ITS dv_promo Non-White Men Force Sust	(3) Logit ITS dv_promo White Women Force Sust	(4) Logit ITS dv_promo Non-White Women Force Sust
Post Policy: post == 1 = 1	0.071**	0.027	-0.003	-0.044
	(0.034)	(0.026)	(0.034)	(0.073)
Age when promoted to Major	-0.002	0.001	0.007	0.006
	(0.001)	(0.002)	(0.008)	(0.005)
Individuals married = 1	0.012	0.007	0.003	-0.049**
	(0.018)	(0.022)	(0.017)	(0.023)
Average Race Similarity to 1LS and 2LS	-0.068	0.025	-0.112**	-0.004
	(0.054)	(0.076)	(0.044)	(0.053)
Average Gender Similarity to 1LS and 2LS	0.064	0.072*	0.363	0.029
	(0.046)	(0.037)	(0.255)	(0.147)
Years of Active Federal Military Service	0.007***	0.004	0.013***	-0.003
	(0.002)	(0.003)	(0.005)	(0.007)
Officer has completed ILE or SSC = 1	0.162***	0.131***		-0.079**
	(0.018)	(0.031)		(0.038)
Any Joint Professional Military Education = 1	-0.035*	-0.012		0.003
	(0.020)	(0.024)		(0.031)
Branch similarity to supervisors	-0.088	-0.039	-0.064	0.047
	(0.057)	(0.040)	(0.069)	(0.043)
Average Senior Rater Score (Scale:1-4)	0.184***	0.197**	0.366***	0.293***
	(0.040)	(0.077)	(0.140)	(0.075)
Average Score of Rater and SR (Scale: 2-8)	0.161***	0.227***	0.029	0.188***
	(0.022)	(0.030)	(0.036)	(0.041)
Year Promoted to Major	-0.021**	-0.004	-0.010	0.008
	(0.010)	(0.008)	(0.011)	(0.022)
Years from first rating officers to eligibility for LTC		-0.019		-0.011
		(0.011)		(0.016)
Years as a supervisor of officers		0.019		-0.024
		(0.017)		(0.027)
Total number of evaluations as supervisor (logged)		0.043		0.080**
		(0.030)		(0.032)
Observations	953	625	164	288
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, by race and gender (Medical Admin and Nurses)

	(1)	(2)	(3)	(4)
	Logit ITS	Logit ITS	Logit ITS	Logit ITS
	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men AMEDD-2	Non-White Men AMEDD-2	White Women AMEDD-2	Non-White Women AMEDD-2
Post Policy: post == 1 = 1	0.095**	0.269**	0.064	0.015
	(0.037)	(0.107)	(0.057)	(0.063)
Age when promoted to Major	0.004*	0.009*	0.002	0.004*
	(0.002)	(0.005)	(0.003)	(0.002)
Individuals married = 1	-0.022	0.032	0.030	0.052
	(0.041)	(0.035)	(0.040)	(0.037)
Average Race Similarity to 1LS and 2LS	-0.065	0.281**	0.173*	-0.127
	(0.060)	(0.112)	(0.097)	(0.115)
Average Gender Similarity to 1LS and 2LS	0.038	0.072	-0.025	0.015
	(0.070)	(0.142)	(0.114)	(0.083)
Years of Active Federal Military Service	0.005*	0.007*	0.005	0.006
	(0.003)	(0.004)	(0.009)	(0.004)
Officer has completed ILE or SSC = 1	0.134***	0.152**	0.095***	0.125***
	(0.026)	(0.065)	(0.032)	(0.042)
Any Joint Professional Military Education = 1	-0.001	0.124*		-0.026
	(0.032)	(0.064)		(0.058)
Branch similarity to supervisors	0.029	-0.027	-0.107	-0.104
	(0.075)	(0.113)	(0.088)	(0.079)
Average Senior Rater Score (Scale:1-4)	0.040	0.128	-0.013	-0.033
	(0.131)	(0.121)	(0.112)	(0.187)
Average Score of Rater and SR (Scale: 2-8)	0.185**	0.143**	0.210***	0.215**
	(0.072)	(0.071)	(0.074)	(0.087)
Year Promoted to Major	-0.034***	-0.079***	-0.032*	-0.001
	(0.010)	(0.030)	(0.019)	(0.013)
Years from first rating officers to eligibility for LTC	0.002	0.062***	0.028*	-0.068***
	(0.014)	(0.015)	(0.016)	(0.019)
Years as a supervisor of officers	0.049*	0.023	0.235***	0.099
	(0.028)	(0.016)	(0.090)	(0.063)
Total number of evaluations as supervisor (logged)	-0.047	-0.122***	-0.304***	0.111***
	(0.031)	(0.045)	(0.109)	(0.041)
Observations	558	280	313	309
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Coarsened Exact Matching

The full coarsened exact matching (CEM) results are presented in Table 2.6. The CEM method is used to balance the covariates between the treatment and control groups, ensuring that the groups are comparable. The table includes the results for the full sample analysis, as well as for the occupational workgroups. The CEM method shows approximately the same results as the primary and linear probability model (LPM) results for the effect of the partial blindness policy on promotion to lieutenant colonel. The following figures show additional results for the four promotion cohorts that officers were considered within.

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Core Employees (CEM)

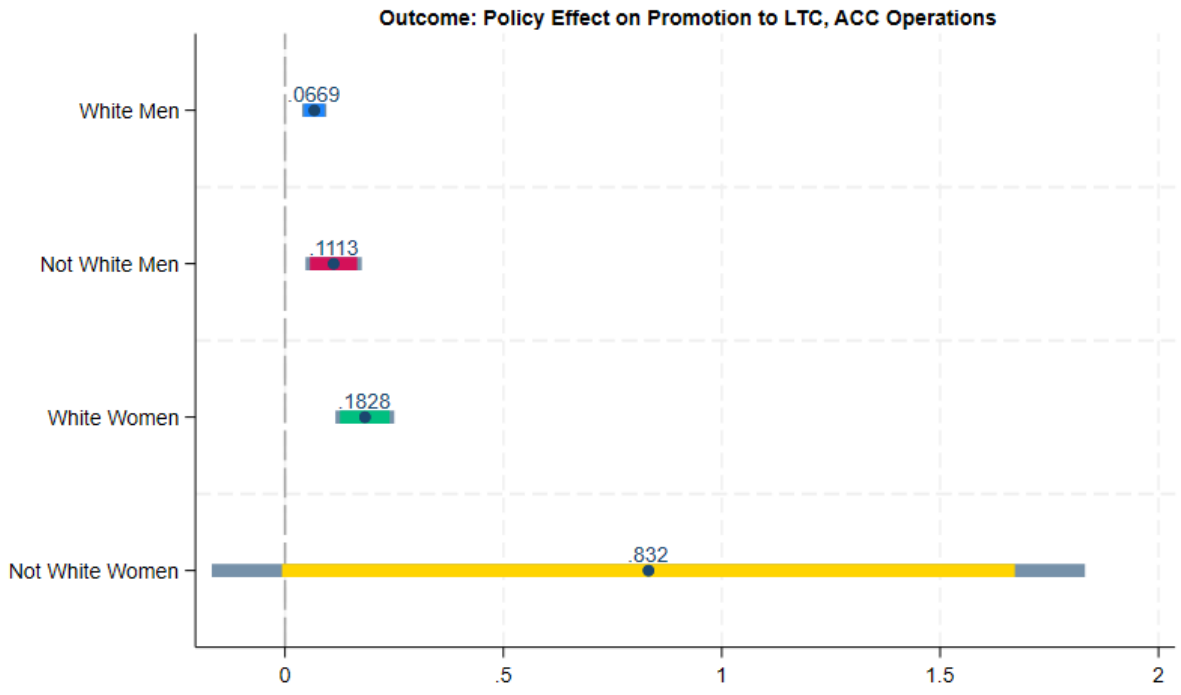

Figure A.1: CEM Model - Core Employees

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Intelligence and Communication Employees (CEM)

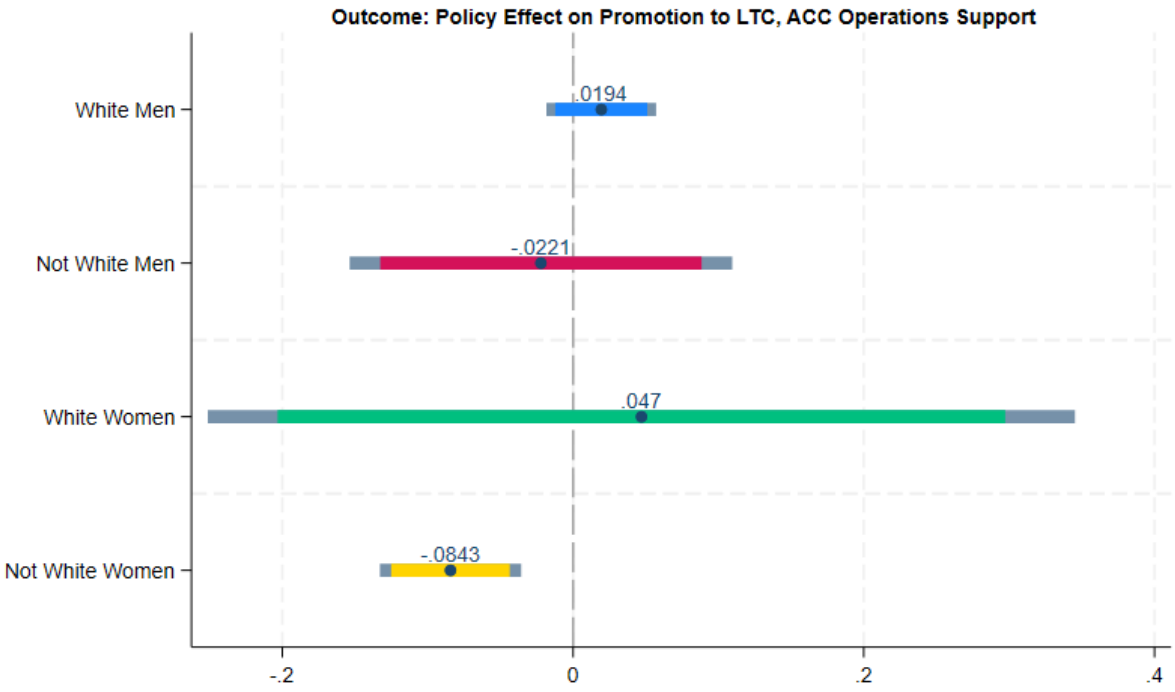

Figure A.2: CEM Model - Intelligence and Communications Employees

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Administrative, Human Resources, and Logistics Employees (CEM)

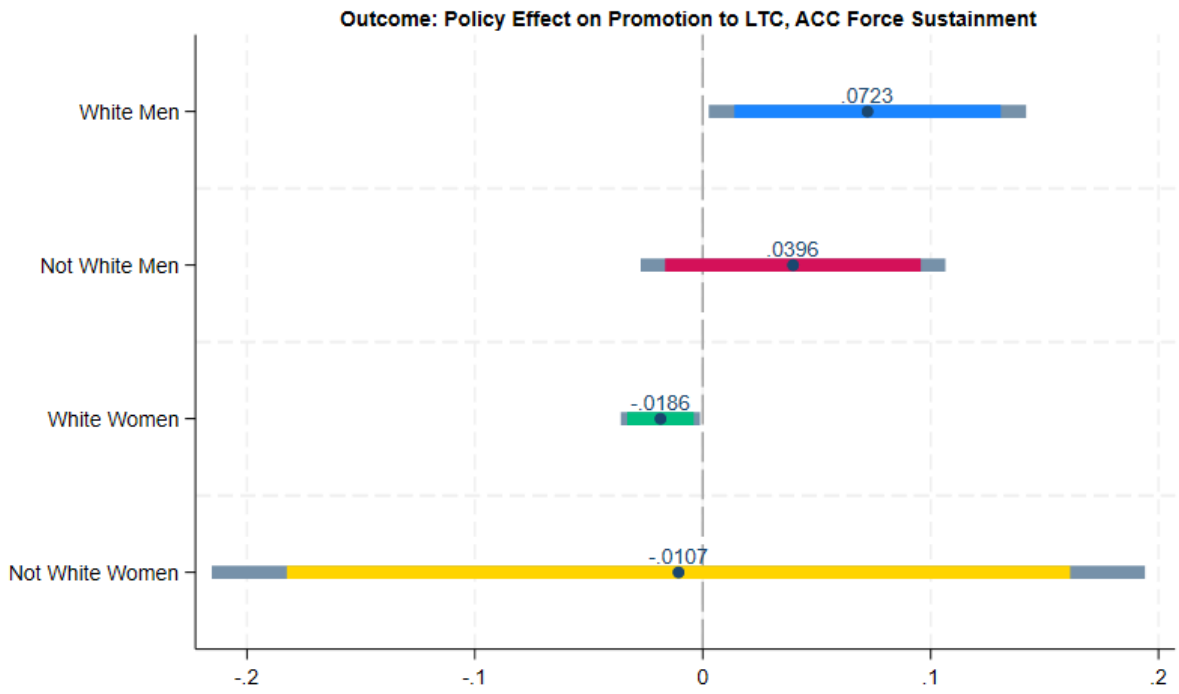

Figure A.3: CEM Model - Administrative, HR, Logistics Employees

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Medical Administrative Employees and Nurses (CEM)

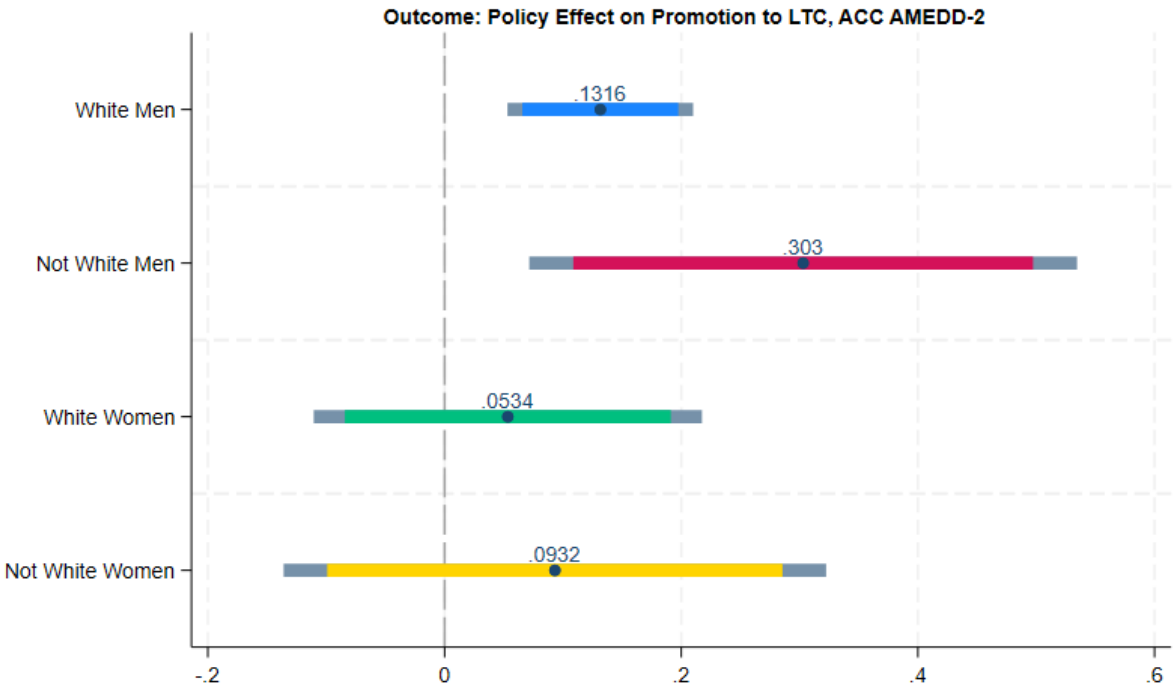

Figure A.4: CEM Model - Medical Administrative Employees and Nurses

Robustness Checks

The robustness checks for the analysis in Chapter 1 are presented in Tables A.5 through A.7. The robustness checks include the full sample analysis, but exclude the average similarity and supervisor measures.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS
	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	Full Sample	All Whites	All Non-Whites	All Men	All Women	White Men	Non-White Men	White Women	Non-White Women
Post Policy: post = 1	0.049***	0.050***	0.045**	0.054***	0.028	0.051***	0.061**	0.071**	-0.006
	(0.011)	(0.012)	(0.022)	(0.013)	(0.021)	(0.013)	(0.027)	(0.029)	(0.038)
All not white individuals = 1	0.001			0.003	-0.014				
	(0.007)			(0.007)	(0.015)				
All females = 1	0.031***	0.038***	0.022						
	(0.008)	(0.010)	(0.015)						
Age when promoted to Major	0.002*	0.001	0.005**	0.001	0.005**	0.000	0.004	0.003	0.006***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)
Individuals married = 1	0.017	0.012	0.019	0.014	0.022	0.017	0.010	0.008	0.027
	(0.011)	(0.014)	(0.016)	(0.014)	(0.017)	(0.016)	(0.020)	(0.019)	(0.023)
Years of Active Federal Military Service	0.004***	0.004***	0.002	0.005***	0.002	0.005***	0.004	0.005	-0.002
	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.005)	(0.004)
Officer has completed ILE or SSC = 1	0.214***	0.229***	0.187***	0.242***	0.151***	0.264***	0.197***	0.095*	0.179***
	(0.022)	(0.029)	(0.025)	(0.023)	(0.036)	(0.028)	(0.029)	(0.049)	(0.054)
Any Joint Professional Military Education = 1	-0.001	-0.003	0.004	-0.006	0.026	-0.008	-0.000	0.038*	0.018
	(0.009)	(0.009)	(0.019)	(0.010)	(0.019)	(0.010)	(0.021)	(0.020)	(0.032)
Command Experience (any rank) = 1	0.023***	0.021**	0.032***	0.021***	0.024*	0.019**	0.032***	0.021**	0.030*
	(0.007)	(0.008)	(0.010)	(0.008)	(0.013)	(0.008)	(0.012)	(0.011)	(0.017)
Average Senior Rater Score (Standardized)	0.023**	0.018	0.033	0.030**	-0.019	0.025**	0.044*	-0.040	-0.003
	(0.010)	(0.012)	(0.021)	(0.011)	(0.019)	(0.013)	(0.025)	(0.031)	(0.035)
Average Composite Score of Rater and SR (Standardized)	0.237***	0.229***	0.249***	0.238***	0.239***	0.232***	0.248***	0.206***	0.261***
	(0.012)	(0.013)	(0.021)	(0.014)	(0.021)	(0.014)	(0.025)	(0.032)	(0.037)
Competitive Category: Operations = 1	0.015	0.024	-0.011	0.012	0.041	0.019	-0.010	0.057	0.004
	(0.028)	(0.022)	(0.050)	(0.028)	(0.041)	(0.022)	(0.051)	(0.036)	(0.077)
Competitive Category: Operations Support = 1	0.030	0.042*	-0.002	0.033	0.015	0.042*	0.010	0.045	-0.025
	(0.028)	(0.024)	(0.051)	(0.028)	(0.040)	(0.024)	(0.051)	(0.040)	(0.074)
Competitive Category: Force Sustainment = 1	0.026	0.041	-0.004	0.028	0.020	0.038	0.001	0.054	-0.010
	(0.029)	(0.025)	(0.049)	(0.030)	(0.039)	(0.026)	(0.050)	(0.035)	(0.072)
Competitive Category: Army Medical Department = 1	0.100***	0.092***	0.100*	0.100**	0.082*	0.099***	0.093	0.053	0.098
	(0.035)	(0.027)	(0.060)	(0.039)	(0.041)	(0.031)	(0.067)	(0.042)	(0.075)
Year Promoted to Major	-0.017***	-0.019***	-0.012*	-0.019***	-0.006	-0.019***	-0.018**	-0.019**	0.006
	(0.003)	(0.003)	(0.007)	(0.003)	(0.006)	(0.003)	(0.007)	(0.009)	(0.009)
Constant	33.826***	37.642***	25.207*	37.923***	13.399	38.337***	37.377**	39.661**	-12.707
	(5.875)	(5.535)	(13.139)	(6.130)	(12.102)	(5.922)	(14.647)	(18.176)	(18.744)
Observations	10,589	7,385	3,204	8,802	1,787	6,464	2,338	921	866
R-squared	0.364	0.345	0.396	0.377	0.319	0.364	0.401	0.224	0.394
ll	-1384	-709.0	-631.9	-1122	-233.4	-630.6	-461.9	-38.06	-161.1
r2_a	0.363	0.344	0.393	0.376	0.314	0.363	0.398	0.212	0.384
rmse	0.276	0.267	0.295	0.275	0.277	0.267	0.296	0.254	0.294
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy
Standard errors in parentheses									
*** p<0.01, ** p<0.05, * p<0.1									

Figure A.5: Linear Probability Model - Full Sample

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS
	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men Operations	Non-White Men Operations	White Women Operations	Non-White Women Operations	White Men Ops Spt	Non-White Men Ops Spt	White Women Ops Spt	Non-White Women Ops Spt
Post Policy: post = 1	0.059***	0.067*	0.079	0.017	-0.003	-0.003	0.093	-0.077
	(0.016)	(0.039)	(0.050)	(0.063)	(0.020)	(0.048)	(0.061)	(0.088)
All not white individuals = 1								
All females = 1								
Age when promoted to Major	-0.002	-0.002	0.001	-0.012	-0.004	0.003	-0.002	0.043***
	(0.002)	(0.005)	(0.018)	(0.019)	(0.003)	(0.007)	(0.004)	(0.009)
Individuals married = 1	0.023	0.012	0.022	0.127*	0.020	-0.012	0.011	0.012
	(0.029)	(0.040)	(0.025)	(0.068)	(0.020)	(0.059)	(0.030)	(0.038)
Years of Active Federal Military Service	0.008***	0.011**	0.007	0.016	0.005	0.002	0.006	-0.042***
	(0.002)	(0.005)	(0.021)	(0.017)	(0.003)	(0.006)	(0.005)	(0.009)
Officer has completed ILE or SSC = 1	0.215***	0.165***	-0.186*	0.262	0.279***	0.205***	0.051	0.269***
	(0.043)	(0.035)	(0.098)	(0.181)	(0.042)	(0.045)	(0.086)	(0.068)
Any Joint Professional Military Education = 1	-0.001	-0.050	-0.046	0.188	0.003	0.018	0.029	0.024
	(0.015)	(0.040)	(0.059)	(0.115)	(0.016)	(0.047)	(0.025)	(0.060)
Command Experience (any rank) = 1	0.027***	0.030	0.043*	0.038	-0.022	0.018	0.038	-0.032
	(0.010)	(0.023)	(0.023)	(0.049)	(0.016)	(0.027)	(0.032)	(0.024)
Average Senior Rater Score (Standardized)	0.030*	0.044	0.037	-0.132**	0.058**	0.104	-0.169***	0.176**
	(0.017)	(0.046)	(0.035)	(0.060)	(0.025)	(0.066)	(0.058)	(0.068)
Average Composite Score of Rater and SR (Standardized)	0.238***	0.282***	0.161**	0.326***	0.230***	0.230***	0.310***	0.252***
	(0.019)	(0.045)	(0.065)	(0.090)	(0.030)	(0.069)	(0.089)	(0.050)
Competitive Category: Operations = 1								
Competitive Category: Operations Support = 1								
Competitive Category: Force Sustainment = 1								
Competitive Category: Army Medical Department = 1								
Year Promoted to Major	-0.021***	-0.025***	-0.012	-0.007	-0.012	-0.003	-0.013	0.015
	(0.003)	(0.008)	(0.013)	(0.009)	(0.010)	(0.016)	(0.013)	(0.022)
Constant	42.829***	50.450***	24.716	13.913	25.583	6.825	26.499	-30.749
	(6.691)	(16.310)	(25.206)	(18.082)	(19.475)	(31.253)	(25.418)	(43.749)
Observations	3,571	917	173	91	990	419	152	147
R-squared	0.361	0.448	0.261	0.384	0.412	0.464	0.257	0.572
ll	-292.7	-153.4	32.43	-14.29	-92.14	-68.39	18.93	-8.810
r2_a	0.360	0.442	0.215	0.307	0.406	0.451	0.205	0.540
rmse	0.263	0.288	0.207	0.302	0.267	0.289	0.222	0.267
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy
Standard errors in parentheses								
*** p<0.01, ** p<0.05, * p<0.1								

Figure A.6: Linear Probability Model - Core Employees, Intelligence, and Communications Employees

	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS	reg ITS
	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men Force Sustainment	Non-White Men Force Sustainment	White Women Force Sustainment	Non-White Women Force Sustainment	White Men AMEDD-2	Non-White Men AMEDD-2	White Women AMEDD-2	Non-White Women AMEDD-2
Post Policy: post = 1	0.062**	0.034	0.050	-0.036	0.082*	0.245**	0.091	0.044
	(0.030)	(0.032)	(0.043)	(0.047)	(0.043)	(0.095)	(0.067)	(0.062)
All not white individuals = 1								
All females = 1								
Age when promoted to Major	-0.001	-0.001	-0.001	0.002	0.005*	0.011**	0.001	0.004
	(0.002)	(0.003)	(0.006)	(0.004)	(0.002)	(0.005)	(0.002)	(0.003)
Individuals married = 1	-0.003	0.019	0.047	-0.039	-0.039	0.044	0.000	0.067
	(0.019)	(0.029)	(0.035)	(0.047)	(0.045)	(0.048)	(0.043)	(0.039)
Years of Active Federal Military Service	0.008***	0.006**	0.008	-0.001	0.006**	0.006	0.005	0.007
	(0.002)	(0.003)	(0.006)	(0.008)	(0.003)	(0.004)	(0.009)	(0.004)
Officer has completed ILE or SSC = 1	0.361***	0.255***	0.009	-0.067	0.284***	0.231**	0.111*	0.227***
	(0.043)	(0.056)	(0.067)	(0.073)	(0.039)	(0.087)	(0.060)	(0.065)
Any Joint Professional Military Education = 1	-0.026*	-0.033	0.041	-0.022	-0.014	0.071**	0.037	-0.027
	(0.015)	(0.030)	(0.035)	(0.057)	(0.019)	(0.032)	(0.028)	(0.071)
Command Experience (any rank) = 1	0.019**	0.013	0.016	0.031**	-0.004	0.055	-0.019	-0.029
	(0.008)	(0.018)	(0.017)	(0.013)	(0.023)	(0.040)	(0.022)	(0.034)
Average Senior Rater Score (Standardized)	0.037*	0.026	0.072*	0.006	-0.038	0.005	-0.069	-0.058
	(0.019)	(0.034)	(0.040)	(0.049)	(0.061)	(0.054)	(0.043)	(0.064)
Average Composite Score of Rater and SR (Standardized)	0.206***	0.270***	0.097**	0.315***	0.198***	0.128*	0.223***	0.225***
	(0.022)	(0.024)	(0.042)	(0.048)	(0.064)	(0.066)	(0.047)	(0.060)
Competitive Category: Operations = 1								
Competitive Category: Operations Support = 1								
Competitive Category: Force Sustainment = 1								
Competitive Category: Army Medical Department = 1								
Year Promoted to Major	-0.016**	-0.002	-0.014	0.016	-0.028**	-0.065**	-0.031	-0.005
	(0.007)	(0.008)	(0.011)	(0.015)	(0.011)	(0.024)	(0.021)	(0.012)
Constant	32.802**	5.073	29.073	-31.694	55.861**	130.956**	63.871	9.631
	(14.881)	(15.696)	(21.791)	(30.820)	(21.359)	(48.495)	(42.844)	(23.297)
Observations	1,042	625	184	288	558	280	338	309
R-squared	0.413	0.461	0.224	0.511	0.288	0.244	0.211	0.341
ll	-10.89	-88.12	35.34	-23.86	-99.82	-69.71	-68.08	-60.29
r2_a	0.408	0.452	0.179	0.493	0.275	0.216	0.187	0.319
rmse	0.246	0.281	0.206	0.268	0.292	0.317	0.301	0.299
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy
Standard errors in parentheses								
*** p<0.01, ** p<0.05, * p<0.1								

Figure A.7: Linear Probability Model - Administrative, HR, Logistics Employees, and Nurses

Relational Demography

The following tables and figures provide additional detail and information of the relational demography analysis for this paper.

Table A.6: Logit Regression Results for Relational Demography Model

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo Greater than mean racial similarity	(3) Logit ITS dv_promo Less than mean racial similarity	(4) Logit ITS dv_promo Greater than mean gender similarity	(5) Logit ITS dv_promo Less than mean gender similarity	(6) Logit ITS dv_promo Greater than mean branch similarity	(7) Logit ITS dv_promo Less than mean branch similarity
Post Policy: post == 1 = 1	0.051***	0.044***	0.059***	0.045***	0.060***	0.060***	0.045***
	(0.011)	(0.011)	(0.018)	(0.014)	(0.014)	(0.013)	(0.014)
All not white individuals = 1	-0.008	0.009	-0.013	0.023**	-0.029*	0.008	-0.014
	(0.010)	(0.021)	(0.011)	(0.010)	(0.015)	(0.014)	(0.013)
All females = 1	0.012	0.037**	-0.003	0.001	0.024	0.022	0.008
	(0.014)	(0.018)	(0.018)	(0.036)	(0.015)	(0.023)	(0.016)
Age when promoted to Major	0.002*	0.001	0.002	0.001	0.002	0.003**	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Individuals married = 1	0.014	0.021	0.008	0.011	0.017	0.027*	0.005
	(0.009)	(0.013)	(0.011)	(0.012)	(0.011)	(0.014)	(0.009)
Average Race Similarity to 1LS and 2LS	-0.015	0.019	-0.005	0.031	-0.051*	0.016	-0.032
	(0.016)	(0.028)	(0.037)	(0.020)	(0.030)	(0.021)	(0.024)
Average Gender Similarity to 1LS and 2LS	-0.027	-0.008	-0.034	-0.016	0.000	-0.021	-0.017
	(0.021)	(0.031)	(0.025)	(0.031)	(0.031)	(0.035)	(0.020)
Years of Active Federal Military Service	0.003***	0.003**	0.004**	0.003**	0.003*	0.004**	0.003**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Officer has completed ILE or SSC = 1	0.126***	0.135***	0.123***	0.143***	0.115***	0.113***	0.140***
	(0.018)	(0.036)	(0.020)	(0.028)	(0.019)	(0.022)	(0.027)
Any Joint Professional Military Education = 1	0.003	0.006	-0.001	-0.008	0.018	0.007	-0.000
	(0.009)	(0.010)	(0.013)	(0.011)	(0.011)	(0.015)	(0.011)
Years from first rating officers to eligibility for LTC	-0.005	-0.003	-0.005	0.002	-0.011**	-0.003	-0.007
	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)	(0.008)	(0.005)
Years as a supervisor of officers	0.025**	0.022	0.030*	0.024*	0.023	0.025**	0.030
	(0.011)	(0.013)	(0.017)	(0.014)	(0.014)	(0.012)	(0.019)
Total number of evaluations as supervisor	-0.002	-0.002	-0.007	-0.014	0.014	-0.006	0.001
	(0.015)	(0.019)	(0.021)	(0.018)	(0.015)	(0.022)	(0.022)
Branch similiarity with supervisors	-0.040**	-0.039*	-0.045**	-0.061**	0.003	-0.014	-0.045
	(0.019)	(0.022)	(0.021)	(0.025)	(0.021)	(0.033)	(0.030)
Average Senior Rater Score (Scale:1-4)	0.184***	0.194***	0.176***	0.220***	0.149***	0.217***	0.167***
	(0.022)	(0.035)	(0.027)	(0.037)	(0.027)	(0.043)	(0.022)
Average Score of Rater and SR (Scale: 2-8)	0.181***	0.174***	0.188***	0.204***	0.150***	0.171***	0.185***
	(0.013)	(0.019)	(0.015)	(0.020)	(0.016)	(0.023)	(0.015)
Competitive Category: Operations = 1	0.011	0.009	0.011	-0.002	0.035	-0.018	0.038
	(0.022)	(0.018)	(0.032)	(0.029)	(0.024)	(0.011)	(0.028)
Competitive Category: Operations Support = 1	0.018	0.023	0.011	0.001	0.038*	0.010	0.029
	(0.021)	(0.018)	(0.031)	(0.029)	(0.022)	(0.013)	(0.026)
Competitive Category: Force Sustainment = 1	0.025	0.035*	0.016	0.031	0.027	0.010	0.041
	(0.021)	(0.019)	(0.031)	(0.029)	(0.022)	(0.015)	(0.025)
Competitive Category: Army Medical Department = 1	0.076***	0.069***	0.081***	0.078***	0.072***	0.064***	0.079***
	(0.018)	(0.015)	(0.027)	(0.025)	(0.020)	(0.014)	(0.020)
Year Promoted to Major	-0.020***	-0.019***	-0.020***	-0.018***	-0.022***	-0.025***	-0.016***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.003)	(0.004)	(0.003)
Constant							
Observations	10,595	5,544	5,051	5,731	4,864	4,730	5,865
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Full Sample

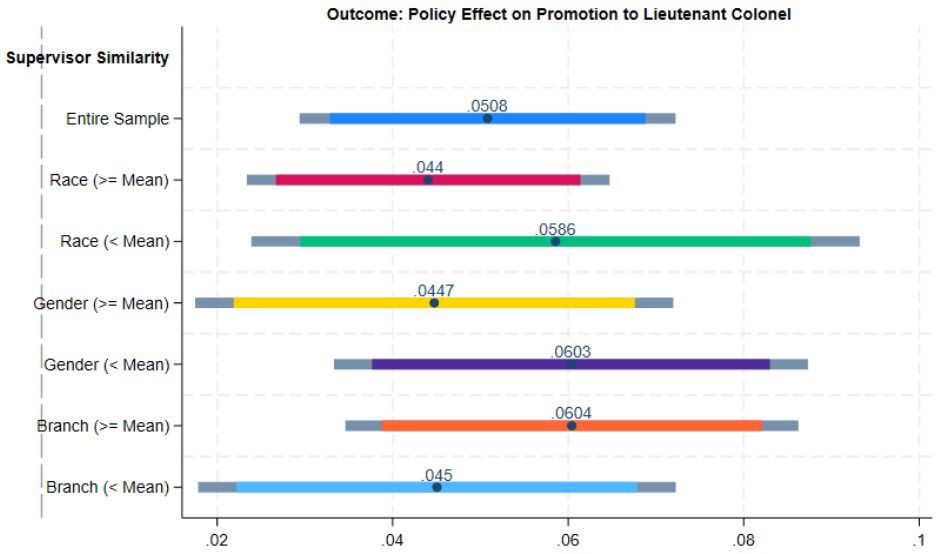

Figure A.8: Primary Results of the Relational Demography Model for Full Sample

Table A.7: Logit Regression Results for Relational Demography Model for Underrepresented Minitorities

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo Greater than mean racial similarity	(3) Logit ITS dv_promo Less than mean racial similarity	(4) Logit ITS dv_promo Greater than mean gender similarity	(5) Logit ITS dv_promo Less than mean gender similarity	(6) Logit ITS dv_promo Greater than mean branch similarity	(7) Logit ITS dv_promo Less than mean branch similarity
Post Policy: post == 1 = 1	0.044**	0.078**	0.040*	0.052*	0.041**	0.083***	0.022
	(0.017)	(0.034)	(0.021)	(0.030)	(0.019)	(0.027)	(0.019)
Age when promoted to Major	0.003**	0.003	0.003*	0.002	0.004**	0.006***	0.001
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Individuals married = 1	0.008	0.026	0.004	-0.001	0.015	0.030*	-0.005
	(0.010)	(0.022)	(0.011)	(0.020)	(0.012)	(0.018)	(0.011)
Years of Active Federal Military Service	0.003*	0.004	0.003	0.005**	0.002	0.004	0.003
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Officer has completed ILE or SSC = 1	0.103***	0.040	0.123***	0.108***	0.101***	0.111***	0.100***
	(0.021)	(0.040)	(0.023)	(0.031)	(0.025)	(0.032)	(0.028)
Any Joint Professional Military Education = 1	0.013	0.010	0.014	-0.034	0.043**	0.002	0.021
	(0.017)	(0.037)	(0.018)	(0.030)	(0.018)	(0.031)	(0.017)
Years from first rating officers to eligibility for LTC	-0.004	-0.004	-0.004	0.008	-0.014	0.003	-0.007
	(0.008)	(0.017)	(0.009)	(0.007)	(0.011)	(0.015)	(0.008)
Years as a supervisor of officers	0.037	-0.005	0.057**	0.006	0.049	0.023	0.075
	(0.025)	(0.021)	(0.026)	(0.039)	(0.032)	(0.028)	(0.049)
n_total_evals_log	-0.017	0.037	-0.041	-0.000	-0.008	-0.008	-0.057
	(0.032)	(0.053)	(0.028)	(0.059)	(0.022)	(0.058)	(0.043)
Average Senior Rater Score (Scale:1-4)	0.185***	0.127	0.196***	0.197***	0.182***	0.231***	0.175***
	(0.042)	(0.084)	(0.047)	(0.067)	(0.049)	(0.050)	(0.053)
Average Score of Rater and SR (Scale: 2-8)	0.211***	0.163***	0.220***	0.266***	0.166***	0.175***	0.219***
	(0.021)	(0.045)	(0.024)	(0.036)	(0.024)	(0.026)	(0.028)
Competitive Category: Operations = 1	-0.005	0.012	-0.009	-0.022	0.011	-0.054*	0.034
	(0.028)	(0.042)	(0.035)	(0.047)	(0.035)	(0.029)	(0.031)
Competitive Category: Operations Support = 1	0.002	0.032	-0.003	-0.021	0.005	-0.025	0.022
	(0.027)	(0.038)	(0.034)	(0.048)	(0.035)	(0.027)	(0.032)
Competitive Category: Force Sustainment = 1	0.012	0.014	0.011	0.025	0.001	-0.031	0.037
	(0.026)	(0.038)	(0.032)	(0.044)	(0.034)	(0.029)	(0.030)
Competitive Category: Army Medical Department = 1	0.073***	0.047	0.079**	0.082**	0.060*	0.042	0.100***
	(0.025)	(0.039)	(0.032)	(0.041)	(0.032)	(0.031)	(0.024)
Year Promoted to Major	-0.015***	-0.025***	-0.013**	-0.016**	-0.015***	-0.029***	-0.006
	(0.005)	(0.008)	(0.006)	(0.008)	(0.006)	(0.007)	(0.006)
Constant							
Observations	4,126	802	3,324	1,548	2,578	1,796	2,330
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Change in Promotion Rates to Lieutenant Colonel by Demographic Group, due the Partial Blindness Policy, Underrepresented Minorities

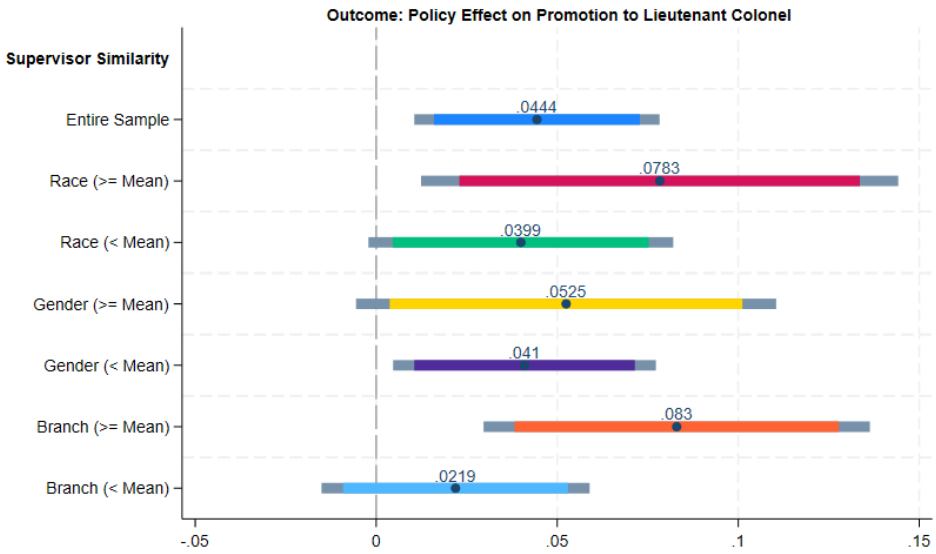

Figure A.9: Primary Results of the Relational Demography Model for Underrepresented Minorities

Appendix B

SENSITIVITY ANALYSIS FOR CHAPTER 1

The sensitivity checks for the analysis in Chapter 1 are presented in Tables B.1 through B.4. The first set of sensitivity checks includes the bottom 50% of performers based on average senior rater score. Tables B.2 through **??** show the results for this analysis. The second set of sensitivity checks excludes promotions to lieutenant colonel in 2021 to account for any potential effects from the COVID-19 pandemic. These results are presented in Tables B.1 through B.4.

Table B.1: Descriptive statistics for bottom 50% of officers

Variables	Count	Mean	SD	Min	Max	Mean (Pre-Policy)	Mean (Post-Policy)	t-statistic	p-value
Promoted to LTC	5208	0.763	0.425	0.000	1.000	0.790	0.732	4.922	0.000
Males promoted to LTC	4261	0.754	0.431	0.000	1.000	0.783	0.717	4.988	0.000
Females promoted to LTC	947	0.808	0.394	0.000	1.000	0.821	0.793	1.082	0.280
Whites promoted to LTC	3325	0.771	0.420	0.000	1.000	0.803	0.733	4.771	0.000
Non-whites promoted to LTC	1883	0.749	0.434	0.000	1.000	0.767	0.730	1.828	0.068
All individuals, exlcuding white males	5208	0.443	0.497	0.000	1.000	0.433	0.455	-1.565	0.118
All not white individuals	5208	0.362	0.480	0.000	1.000	0.352	0.373	-1.566	0.118
All females	5208	0.182	0.386	0.000	1.000	0.174	0.191	-1.551	0.121
Age when promoted to Major	5208	38.315	4.280	30.000	57.000	38.134	38.530	-3.330	0.001
Individuals married	5208	0.855	0.352	0.000	1.000	0.859	0.851	0.858	0.391
Employee and supervisor are same gender (binary)	5208	0.506	0.500	0.000	1.000	0.526	0.482	3.158	0.002
Employee and supervisor are same race (binary)	5208	0.391	0.488	0.000	1.000	0.414	0.364	3.652	0.000
Employee and 2LS are same gender (binary)	5208	0.535	0.499	0.000	1.000	0.557	0.508	3.539	0.000
Employee and 2LS are same race (binary)	5208	0.415	0.493	0.000	1.000	0.428	0.398	2.188	0.029
Employee and 1LS are same occupational workgroup	5208	0.287	0.452	0.000	1.000	0.288	0.286	0.206	0.837
Employee and 2LS are same occupational workgroup	5208	0.223	0.416	0.000	1.000	0.234	0.209	2.139	0.032
Years of AFMS when eligible for promotion to LTC	5208	19.800	3.432	6.000	37.000	19.633	19.997	-3.821	0.000
Officer has completed ILE or SSC	5208	0.936	0.244	0.000	1.000	0.934	0.939	-0.794	0.427
Any Joint Professional Military Education	5208	0.092	0.290	0.000	1.000	0.098	0.085	1.651	0.099
Years from first rating officers to eligibility for LTC	5208	0.531	1.965	0.000	17.000	0.367	0.724	-6.554	0.000
Years as a supervisor of officers	5208	0.442	1.644	0.000	12.000	0.386	0.508	-2.669	0.008
Number of officer evaluations as a supervisor	5208	5.674	25.740	0.000	568.000	4.699	6.829	-2.977	0.003
Average Senior Rater Score (Scale:1-4)	5208	3.405	0.187	3.000	3.652	3.402	3.409	-1.258	0.208
Average Score of Rater and SR (Scale: 2-8)	5208	7.081	0.378	5.667	7.652	7.149	7.001	14.333	0.000

Note: The table presents descriptive statistics for the full sample (N=10,595). The group means are identified by if the employees primary promotion window occurred before or after the partial blindness policy went into effect. AFMS = Active Federal Military Service. SD = Standard Deviation. 1LS = First Line Supervisor (Rater). 2LS = Second Line Supervisor (Senior Rater).

Table B.2: Logit Regression Results for Promotion to Lieutenant Colonel (Bottom 50% by Performance)

VARIABLES	(1) Logit ITS dv_promo Full Sample	(2) Logit ITS dv_promo All Whites	(3) Logit ITS dv_promo All Non-Whites	(4) Logit ITS dv_promo All Men	(5) Logit ITS dv_promo All Women
Post Policy: post == 1 = 1	0.042**	0.045**	0.028	0.052**	-0.001
	(0.019)	(0.021)	(0.034)	(0.022)	(0.037)
All not white individuals = 1	-0.018			-0.005	-0.046
	(0.017)			(0.016)	(0.040)
All females = 1	-0.002	0.020	-0.012		
	(0.024)	(0.031)	(0.032)		
Age when promoted to Major	0.001	-0.000	0.003	-0.000	0.004
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Individuals married = 1	0.029**	0.032*	0.022	0.021	0.040
	(0.015)	(0.018)	(0.019)	(0.014)	(0.025)
Average Race Similarity to 1LS and 2LS	-0.046	-0.027	-0.000	-0.031	-0.047
	(0.028)	(0.032)	(0.056)	(0.030)	(0.055)
Average Gender Similarity to 1LS and 2LS	-0.068*	-0.102**	-0.036	-0.089**	0.014
	(0.036)	(0.044)	(0.044)	(0.041)	(0.076)
Years of Active Federal Military Service	0.004*	0.004*	0.003	0.004*	0.005
	(0.002)	(0.002)	(0.003)	(0.002)	(0.005)
Officer has completed ILE or SSC = 1	0.172***	0.141***	0.146***	0.159***	0.103***
	(0.029)	(0.029)	(0.026)	(0.024)	(0.035)
Any Joint Professional Military Education = 1	0.014	-0.006	0.057**	0.005	0.061
	(0.015)	(0.019)	(0.023)	(0.016)	(0.043)
Years from first rating officers to eligibility for LTC	-0.006	0.007	-0.020**	-0.004	0.000
	(0.007)	(0.009)	(0.010)	(0.006)	(0.024)
Years as a supervisor of officers	0.045**	0.036**	0.066**	0.036*	0.079**
	(0.018)	(0.018)	(0.033)	(0.022)	(0.040)
Total evals as 1LS (logged)	-0.014	-0.031	0.004	-0.004	-0.077
	(0.027)	(0.026)	(0.036)	(0.030)	(0.085)
Average Branch Similarity to 1LS and 2LS	-0.062**	-0.060*	-0.075*	-0.063*	-0.049
	(0.030)	(0.034)	(0.040)	(0.035)	(0.044)
Average Senior Rater Score (Scale:1-4)	0.387***	0.351***	0.472***	0.418***	0.263***
	(0.038)	(0.046)	(0.074)	(0.043)	(0.079)
Average Score of Rater and SR (Scale: 2-8)	0.310***	0.316***	0.291***	0.308***	0.309***
	(0.024)	(0.027)	(0.039)	(0.027)	(0.038)
Competitive Category: Operations = 1	0.002	0.019	-0.031	0.006	0.026
	(0.037)	(0.030)	(0.053)	(0.039)	(0.054)
Competitive Category: Operations Support = 1	0.003	0.026	-0.043	0.017	-0.040
	(0.037)	(0.034)	(0.053)	(0.039)	(0.046)
Competitive Category: Force Sustainment = 1	0.034	0.067*	-0.011	0.046	0.009
	(0.036)	(0.036)	(0.051)	(0.041)	(0.044)
Competitive Category: Army Medical Department = 1	0.146***	0.151***	0.173***	0.178***	0.120***
	(0.032)	(0.039)	(0.062)	(0.053)	(0.044)
Year Promoted to Major	-0.014***	-0.017***	-0.006	-0.018***	0.004
	(0.005)	(0.005)	(0.009)	(0.005)	(0.011)
Observations	5,208	3,325	1,883	4,261	947
Cluster	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
WM = White Men, NWM = Non-White Men, WW = White Women, NWW = Non-White Women
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.3: Logit Regression Results for Promotion to Lieutenant Colonel, by Race and Gender (Bottom 50% by Performance)

VARIABLES	(1) Logit ITS dv_promo White Men	(2) Logit ITS dv_promo Non-White Men	(3) Logit ITS dv_promo White Women	(4) Logit ITS dv_promo Non-White Women
Post Policy: post == 1 = 1	0.046**	0.057	0.044	-0.055
	(0.023)	(0.038)	(0.043)	(0.062)
Age when promoted to Major	-0.001	0.001	0.004	0.007*
	(0.002)	(0.003)	(0.003)	(0.004)
Individuals married = 1	0.025	0.013	0.053	0.031
	(0.016)	(0.023)	(0.034)	(0.027)
Average Race Similarity to 1LS and 2LS	-0.017	0.046	0.010	-0.124
	(0.040)	(0.071)	(0.067)	(0.097)
Average Gender Similarity to 1LS and 2LS	-0.124**	-0.047	0.060	-0.015
	(0.052)	(0.053)	(0.095)	(0.114)
Years of Active Federal Military Service	0.004	0.004	0.012*	-0.001
	(0.002)	(0.003)	(0.007)	(0.005)
Officer has completed ILE or SSC = 1	0.163***	0.157***	0.084**	0.123**
	(0.034)	(0.030)	(0.042)	(0.051)
Any Joint Professional Military Education = 1	-0.012	0.051**	0.105	0.053
	(0.019)	(0.026)	(0.097)	(0.053)
Years from first rating officers to eligibility for LTC	0.003	-0.011	0.052	-0.057**
	(0.009)	(0.008)	(0.033)	(0.028)
Years as a supervisor of officers	0.033	0.031	0.107*	0.072
	(0.021)	(0.040)	(0.065)	(0.046)
Total evals as 1LS (logged)	-0.015	0.026	-0.193**	0.102
	(0.030)	(0.056)	(0.091)	(0.078)
Average Branch Similarity to 1LS and 2LS	-0.061	-0.081*	-0.090	-0.041
	(0.038)	(0.046)	(0.066)	(0.063)
Average Senior Rater Score (Scale:1-4)	0.393***	0.506***	0.094	0.361***
	(0.049)	(0.090)	(0.136)	(0.140)
Average Score of Rater and SR (Scale: 2-8)	0.313***	0.284***	0.300***	0.317***
	(0.028)	(0.046)	(0.064)	(0.069)
Competitive Category: Operations = 1	0.029	-0.051	-0.003	0.048
	(0.028)	(0.061)	(0.059)	(0.062)
Competitive Category: Operations Support = 1	0.037	-0.038	0.006	-0.053
	(0.030)	(0.061)	(0.071)	(0.050)
Competitive Category: Force Sustainment = 1	0.081**	-0.026	0.012	0.031
	(0.034)	(0.059)	(0.065)	(0.045)
Competitive Category: Army Medical Department = 1	0.171***	0.162**	0.047	0.186***
	(0.045)	(0.074)	(0.048)	(0.056)
Year Promoted to Major	-0.019***	-0.015	-0.011	0.019
	(0.006)	(0.010)	(0.014)	(0.015)
Observations	2,901	1,360	424	523
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
WM = White Men, NWM = Non-White Men, WW = White Women, NWW = Non-White Women
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.4: Logit Regression Results for Promotion to LTC by Core Operations Employees (Bottom 50% by Performance)

VARIABLES	(10) Logit ITS dv_promo WM Operations	(11) Logit ITS dv_promo NWM Operations	(12) Logit ITS dv_promo WW Operations	(13) Logit ITS dv_promo NWM Operations
Post Policy: post == 1 = 1	0.051	0.089**	-0.000	-0.000
	(0.035)	(0.042)	(0.000)	(0.000)
Age when promoted to Major	-0.002	-0.001	-0.000	0.000
	(0.003)	(0.005)	(0.000)	(0.000)
Individuals married = 1	0.039	0.023	0.000	0.000
	(0.024)	(0.045)	(0.000)	(0.000)
Average Race Similarity to 1LS and 2LS	-0.026	-0.030	-0.000	0.000
	(0.053)	(0.128)	(0.000)	(0.000)
Average Gender Similarity to 1LS and 2LS	-0.100	0.014	0.000	-0.000
	(0.068)	(0.081)	(0.000)	(0.000)
Years of Active Federal Military Service	0.002	0.008	0.000	0.000
	(0.004)	(0.007)	(0.000)	(0.000)
Officer has completed ILE or SSC = 1	0.170**	0.143***		
	(0.068)	(0.028)		
Any Joint Professional Military Education = 1	0.007	0.026	-0.000	
	(0.025)	(0.043)	(0.000)	
Years from first rating officers to eligibility for LTC	-0.023	-0.011		
	(0.017)	(0.018)		
Years as a supervisor of officers	0.091**	0.197***		
	(0.036)	(0.040)		
Total evals as 1LS (logged)	-0.038	-0.110**		
	(0.037)	(0.053)		
Average Branch Similarity to 1LS and 2LS	-0.132***	-0.217**	-0.000	-0.000
	(0.042)	(0.088)	(0.000)	(0.000)
Average Senior Rater Score (Scale:1-4)	0.410***	0.617***	0.000	-0.000
	(0.069)	(0.134)	(0.000)	(0.000)
Average Score of Rater and SR (Scale: 2-8)	0.326***	0.268***	0.000	0.000
	(0.043)	(0.066)	(0.000)	(0.000)
Observations	1,493	504	54	42
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
WM = White Men, NWM = Non-White Men, WW = White Women, NWW = Non-White Women
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.5: Logit Regression Results for Promotion to LTC by Intelligence and Communications Employees (Bottom 50% by Performance)

VARIABLES	(1) Logit ITS dv_promo WM Ops Spt	(2) Logit ITS dv_promo NWM Ops Spt	(3) Logit ITS dv_promo WW Ops Spt	(4) Logit ITS dv_promo NWW Ops Spt
Post Policy: post == 1 = 1	-0.034	-0.053	-0.000	-0.116**
	(0.035)	(0.075)	(0.000)	(0.056)
Age when promoted to Major	-0.006	-0.004	-0.000	0.053***
	(0.004)	(0.010)	(0.000)	(0.010)
Individuals married = 1	0.006	-0.005	0.000	0.116**
	(0.034)	(0.075)	(0.000)	(0.047)
Average Race Similarity to 1LS and 2LS	0.075	0.074	0.000	0.178**
	(0.139)	(0.100)	(0.000)	(0.076)
Average Gender Similarity to 1LS and 2LS	-0.271*	-0.254***	-0.000	-1.384***
	(0.150)	(0.086)	(0.000)	(0.408)
Years of Active Federal Military Service	-0.001	0.012	0.000	-0.038***
	(0.005)	(0.013)	(0.000)	(0.012)
Officer has completed ILE or SSC = 1	0.172***	0.075	0.000	
	(0.059)	(0.051)	(0.000)	
Any Joint Professional Military Education = 1	0.037	0.103		-0.170***
	(0.083)	(0.073)		(0.063)
Years from first rating officers to eligibility for LTC	0.028*	-0.002		
	(0.016)	(0.007)		
Years as a supervisor of officers	0.031	-0.022		
	(0.028)	(0.064)		
Total evals as 1LS (logged)	-0.075	0.093		
	(0.055)	(0.099)		
Average Branch Similarity to 1LS and 2LS	-0.016	-0.060	0.000	0.197***
	(0.055)	(0.049)	(0.000)	(0.076)
Average Senior Rater Score (Scale:1-4)	0.413***	0.768***	-0.000	0.588***
	(0.130)	(0.293)	(0.000)	(0.211)
Average Score of Rater and SR (Scale: 2-8)	0.319***	0.183	0.000	0.613***
	(0.064)	(0.171)	(0.000)	(0.058)
Observations	486	243	62	76
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
WM = White Men, NWM = Non-White Men, WW = White Women, NWW = Non-White Women
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.6: Logit Regression Results for Promotion to LTC by Administrative, HR, and Logistics Employees

VARIABLES	(18) Logit ITS dv_promo WM Force Sustianment	(19) Logit ITS dv_promo NWM Force Sustainment	(20) Logit ITS dv_promo WW Force Sustainment	(21) Logit ITS dv_promo NWW Force Sustainment
Post Policy: post == 1 = 1	0.082	0.004	0.000	-0.106
	(0.063)	(0.052)	(0.000)	(0.121)
Age when promoted to Major	-0.002	-0.001	0.000	0.008
	(0.004)	(0.003)	(0.000)	(0.009)
Individuals married = 1	0.031	0.020	0.000	-0.086**
	(0.023)	(0.038)	(0.000)	(0.040)
Average Race Similarity to 1LS and 2LS	-0.267***	0.069	0.000	-0.008
	(0.091)	(0.121)	(0.000)	(0.110)
Average Gender Similarity to 1LS and 2LS	0.166*	0.112**	0.000	-0.025
	(0.087)	(0.056)	(0.000)	(0.227)
Years of Active Federal Military Service	0.009**	0.005	-0.000	-0.006
	(0.004)	(0.005)	(0.000)	(0.012)
Officer has completed ILE or SSC = 1	0.275***	0.226***		-0.125**
	(0.022)	(0.052)		(0.058)
Any Joint Professional Military Education = 1	-0.078***	0.011		0.022
	(0.029)	(0.025)		(0.063)
Years from first rating officers to eligibility for LTC		-0.031**		-0.017
		(0.015)		(0.036)
Years as a supervisor of officers		0.020		-0.046
		(0.024)		(0.044)
Total evals as 1LS (logged)		0.073*		0.135
		(0.042)		(0.086)
Average Branch Similarity to 1LS and 2LS	-0.148**	-0.073	-0.000	0.066
	(0.072)	(0.062)	(0.000)	(0.074)
Average Senior Rater Score (Scale:1-4)	0.455***	0.367***	0.000	0.510***
	(0.066)	(0.104)	(0.000)	(0.149)
Average Score of Rater and SR (Scale: 2-8)	0.302***	0.363***	0.000	0.336***
	(0.039)	(0.042)	(0.000)	(0.081)
Observations	424	361	59	158
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
WM = White Men, NWM = Non-White Men, WW = White Women, NWW = Non-White Women
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B.7: Logit Regression Results for Promotion to LTC by Medical Administrative Employees and Nurses

VARIABLES	(22) Logit ITS dv_promo WM AMEDD-2	(23) Logit ITS dv_promo NWM AMEDD-2	(24) Logit ITS dv_promo WW AMEDD-2	(25) Logit ITS dv_promo NWW AMEDD-2
Post Policy: post == 1 = 1	0.096	0.290***	0.057	-0.046
	(0.064)	(0.108)	(0.086)	(0.100)
Age when promoted to Major	0.007*	0.010***	0.002	0.005
	(0.004)	(0.004)	(0.003)	(0.004)
Individuals married = 1	-0.009	0.041	0.090	0.043
	(0.054)	(0.045)	(0.064)	(0.040)
Average Race Similarity to 1LS and 2LS	-0.125	0.404**	0.117	-0.238**
	(0.081)	(0.173)	(0.141)	(0.112)
Average Gender Similarity to 1LS and 2LS	0.004	0.009	-0.003	0.087
	(0.106)	(0.185)	(0.201)	(0.082)
Years of Active Federal Military Service	0.006	0.003	0.014	0.006
	(0.004)	(0.004)	(0.011)	(0.006)
Officer has completed ILE or SSC = 1	0.159***	0.157***	0.134***	0.149***
	(0.034)	(0.054)	(0.042)	(0.041)
Any Joint Professional Military Education = 1	-0.018	0.083***		0.032
	(0.044)	(0.028)		(0.078)
Years from first rating officers to eligibility for LTC	0.039	0.084***	0.063**	-0.072***
	(0.025)	(0.021)	(0.030)	(0.028)
Years as a supervisor of officers	0.044*	0.074***	0.221***	0.092*
	(0.027)	(0.015)	(0.079)	(0.054)
Total evals as 1LS (logged)	-0.091*	-0.199***	-0.339***	0.125*
	(0.055)	(0.041)	(0.110)	(0.075)
Average Branch Similarity to 1LS and 2LS	0.114	-0.028	-0.097	-0.150
	(0.094)	(0.088)	(0.142)	(0.094)
Average Senior Rater Score (Scale:1-4)	0.190	0.383*	0.118	0.128
	(0.162)	(0.202)	(0.212)	(0.250)
Average Score of Rater and SR (Scale: 2-8)	0.250***	0.184**	0.215**	0.177
	(0.096)	(0.088)	(0.095)	(0.112)
Observations	314	196	185	210
Cluster	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

1LS or Rater = First line supervisor. 2LS or SR = 2nd Line Supervisor. ILE = Intermediate Level Education, a required professional school for majors
WM = White Men, NWM = Non-White Men, WW = White Women, NWW = Non-White Women
LTC = Lieutenant Colonel. SSC = Senior Service College - a higher level professional schooling for promotable lieutenant colonels and colonels.
Competitive Category: Operations = Occupational Workgroup - Core Employees.
Competitive Category: Operations Support = Occupational Workgroup - Intelligence and Communications Employees.
Competitive Category: Force Sustainment = Occupational Workgroup - Administrative, Human Resources, and Logistics Employees.
Competitive Category: Army Medical Department = Occupational Workgroup - Medical Administration and Nurses.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix B: Descriptive Statistics

	Full Sample							
	Count	Mean	SD	Min	Max	Mean (Pre-Policy)	Mean (Post-Policy)	t-statistic	p-value
Promoted to LTC	7806	0.855	0.352	0.000	1.000	0.870	0.853	0.961	0.337
Males promoted to LTC	6466	0.853	0.354	0.000	1.000	0.870	0.839	1.500	0.134
Females promoted to LTC	1340	0.867	0.340	0.000	1.000	0.873	0.901	-0.803	0.423
Whites promoted to LTC	5426	0.871	0.336	0.000	1.000	0.917	0.865	2.524	0.012
Non-whites promoted to LTC	2380	0.820	0.384	0.000	1.000	0.807	0.838	-1.063	0.288
All individuals, excluding white males	7806	0.392	0.488	0.000	1.000	0.502	0.523	-0.836	0.403
All not white individuals	7806	0.305	0.460	0.000	1.000	0.421	0.427	-0.214	0.831
All females	7806	0.172	0.377	0.000	1.000	0.208	0.231	-1.119	0.263
Age when promoted to Major	7806	37.584	3.999	30.000	57.000	37.711	38.272	-2.910	0.004
Individuals married	7806	0.880	0.325	0.000	1.000	0.839	0.865	-1.417	0.157
Years of AFMS when eligible for promotion to LTC	7806	19.608	3.216	6.000	37.000	19.986	20.503	-2.904	0.004
Officer has completed ILE or SSC	7806	0.959	0.199	0.000	1.000	0.966	0.984	-2.299	0.022
Any Joint Professional Military Education	7806	0.104	0.306	0.000	1.000	0.083	0.111	-1.876	0.061
Command Experience (any rank)	7806	0.236	0.424	0.000	1.000	0.293	0.300	-0.324	0.746
Average Senior Rater Score (Scale:1-4)	7806	3.623	0.278	2.355	4.000	3.604	3.636	-2.178	0.030
Average Score of Rater and SR (Scale: 2-8)	7806	7.373	0.463	4.968	8.000	7.397	7.336	2.482	0.013

Figure B.1: Descriptive Statistics – Promotions from 2021 removed

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS
	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	Full Sample	All Whites	All Non-Whites	All Men	All Women	White Men	Non-White Men	White Women	Non-White Women
Post Policy: post == 1 = 1	0.087***	0.082***	0.087**	0.092***	0.073**	0.090***	0.095**	0.063	0.839
	(0.017)	(0.019)	(0.039)	(0.019)	(0.032)	(0.020)	(0.040)	(0.060)	(0.705)
All not white individuals = 1	-0.002			-0.003	0.006				
	(0.006)			(0.006)	(0.020)				
All females = 1	0.029***	0.029**	0.034**						
	(0.009)	(0.014)	(0.015)						
Age when promoted to Major	0.002	0.001	0.003	0.001	0.003	0.001	0.004	0.004	0.031
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.003)	(0.039)
Individuals married = 1	0.016	0.015	0.014	0.012	0.022	0.013	0.007	0.022	0.136
	(0.011)	(0.012)	(0.014)	(0.011)	(0.020)	(0.012)	(0.018)	(0.026)	(0.296)
Years of Active Federal Military Service	0.005***	0.005***	0.005**	0.006***	0.003	0.005***	0.008***	0.010	-0.014
	(0.001)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.007)	(0.055)
Officer has completed ILE or SSC = 1	0.115***	0.086***	0.102***	0.099***	0.092***	0.100***	0.099***	0.062	1.393***
	(0.020)	(0.019)	(0.022)	(0.016)	(0.025)	(0.021)	(0.024)	(0.042)	(0.400)
Any Joint Professional Military Education = 1	-0.007	-0.012	0.006	-0.015	0.080	-0.018	-0.005	0.129	0.716
	(0.012)	(0.013)	(0.021)	(0.012)	(0.054)	(0.013)	(0.024)	(0.097)	(0.759)
Command Experience (any rank) = 1	0.131***	0.194***	0.293***	0.203***		0.182***	0.276***		
	(0.007)	(0.030)	(0.086)	(0.025)		(0.029)	(0.084)		
Average Senior Rater Score (Scale:1-4)	0.124***	0.113***	0.143**	0.137***	0.068	0.127***	0.156**	0.027	1.389
	(0.023)	(0.028)	(0.061)	(0.026)	(0.062)	(0.031)	(0.073)	(0.092)	(1.531)
Average Score of Rater and SR (Scale: 2-8)	0.203***	0.189***	0.237***	0.202***	0.255***	0.188***	0.243***	0.223***	2.976***
	(0.012)	(0.014)	(0.030)	(0.014)	(0.027)	(0.015)	(0.036)	(0.049)	(0.730)
Competitive Category: Operations = 1	0.004	0.010	-0.017	0.003	0.030	0.012	-0.037	0.041	0.247
	(0.021)	(0.016)	(0.037)	(0.020)	(0.051)	(0.015)	(0.035)	(0.057)	(0.662)
Competitive Category: Operations Support = 1	0.022	0.036**	-0.014	0.029	-0.002	0.042**	-0.020	0.024	-0.202
	(0.019)	(0.017)	(0.037)	(0.020)	(0.043)	(0.016)	(0.034)	(0.050)	(0.601)
Competitive Category: Force Sustainment = 1	0.022	0.036*	-0.007	0.025	0.026	0.041**	-0.026	0.049	0.245
	(0.020)	(0.020)	(0.035)	(0.021)	(0.043)	(0.020)	(0.032)	(0.047)	(0.559)
Competitive Category: Army Medical Department = 1	0.067***	0.065***	0.081*	0.075***	0.087**	0.078***	0.042	0.039	1.523***
	(0.019)	(0.019)	(0.046)	(0.028)	(0.044)	(0.021)	(0.050)	(0.053)	(0.591)
Year Promoted to Major	-0.032***	-0.033***	-0.029***	-0.035***	-0.022**	-0.035***	-0.034***	-0.020	-0.229
	(0.004)	(0.005)	(0.010)	(0.005)	(0.010)	(0.005)	(0.011)	(0.016)	(0.180)
Observations	7,806	5,426	2,380	6,466	1,083	4,748	1,718	535	548
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy
Standard errors in parentheses									
*** p<0.01, ** p<0.05, * p<0.1									

Figure B.2: Sensitivity Check Results for full panel, excluding promotions to lieutenant colonel in 2021

	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS
	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men Operations	Non-White Men Operations	White Women Operations	Non-White Women Operations	White Men Ops Spt	Non-White Men Ops Spt	White Women Ops Spt
Post Policy: post == 1 = 1	0.099***	0.143***	0.030	0.163	0.031	-0.099	0.127
	(0.023)	(0.044)	(0.061)	(0.184)	(0.032)	(0.092)	(0.152)
All not white individuals = 1							
All females = 1							
Age when promoted to Major	-0.000	0.001	-0.011	-0.014	-0.002	-0.005	-0.021**
	(0.001)	(0.005)	(0.013)	(0.015)	(0.003)	(0.007)	(0.010)
Individuals married = 1	0.017	0.013	0.022	0.181*	-0.004	0.042	0.061
	(0.019)	(0.025)	(0.029)	(0.099)	(0.023)	(0.069)	(0.064)
Years of Active Federal Military Service	0.006**	0.014***	0.029	0.037	0.003	0.010	0.023
	(0.002)	(0.005)	(0.032)	(0.023)	(0.004)	(0.011)	(0.026)
Officer has completed ILE or SSC = 1	0.066**	0.102**		0.249**	0.130***	0.023	-0.014
	(0.030)	(0.040)		(0.121)	(0.041)	(0.037)	(0.105)
Any Joint Professional Military Education = 1	-0.002	-0.029	-0.168***		-0.032	0.031	
	(0.023)	(0.040)	(0.048)		(0.022)	(0.083)	
Command Experience (any rank) = 1	0.162***	0.145***					
	(0.032)	(0.054)					
Average Senior Rater Score (Scale:1-4)	0.125***	0.180	0.258*	-0.608**	0.232**	0.273*	-0.465
	(0.035)	(0.141)	(0.137)	(0.243)	(0.104)	(0.142)	(0.384)
Average Score of Rater and SR (Scale: 2-8)	0.192***	0.248***	0.134*	0.708***	0.205***	0.307***	0.483***
	(0.019)	(0.072)	(0.070)	(0.128)	(0.050)	(0.072)	(0.151)
Competitive Category: Operations = 1							
Competitive Category: Operations Support = 1							
Competitive Category: Force Sustainment = 1							
Competitive Category: Army Medical Department = 1							
Year Promoted to Major	-0.038***	-0.052***	-0.007	-0.052	-0.025**	0.012	-0.019
	(0.007)	(0.013)	(0.025)	(0.045)	(0.011)	(0.018)	(0.036)
Observations	2,614	673	79	54	615	247	74
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy
Standard errors in parentheses							
*** p<0.01, ** p<0.05, * p<0.1							

Figure B.3: Sensitivity Check Results for full panel, excluding promotions to lieutenant colonel in 2021, Core Employees, Intelligence and Communications Occupational Groups

	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS	Logit ITS
	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo	dv_promo
VARIABLES	White Men Force Sustainment	Non-White Men Force Sustainment	White Women Force Sustainment	Non-White Women Force Sustainment	White Men AMEDD-2	Non-White Men AMEDD-2	White Women AMEDD-2	Non-White Women AMEDD-2
Post Policy: post == 1 = 1	0.189***	-0.000	0.062	0.019	0.117	0.544***	0.033	0.179
	(0.052)	(0.040)	(0.059)	(0.125)	(0.076)	(0.180)	(0.130)	(0.114)
All not white individuals = 1								
All females = 1								
Age when promoted to Major	-0.001	-0.005*	0.006	0.001	0.004	0.013*	0.003	-0.001
	(0.002)	(0.003)	(0.009)	(0.007)	(0.003)	(0.008)	(0.004)	(0.003)
Individuals married = 1	0.015	0.013	-0.029	-0.070**	-0.014	0.005	0.028	0.058
	(0.018)	(0.033)	(0.029)	(0.032)	(0.060)	(0.107)	(0.045)	(0.043)
Years of Active Federal Military Service	0.012***	0.013***	0.015	-0.003	0.010***	0.005	0.010	0.007*
	(0.004)	(0.003)	(0.011)	(0.010)	(0.003)	(0.006)	(0.012)	(0.004)
Officer has completed ILE or SSC = 1	0.299***	0.213***		-0.069	0.104***	0.196**	0.113**	0.144***
	(0.050)	(0.043)		(0.055)	(0.027)	(0.077)	(0.049)	(0.039)
Any Joint Professional Military Education = 1	-0.066***	-0.040		0.065	-0.028			-0.005
	(0.023)	(0.059)		(0.081)	(0.039)			(0.088)
Command Experience (any rank) = 1					0.101*			
					(0.054)			
Average Senior Rater Score (Scale:1-4)	0.226***	0.100	0.375***	0.366***	-0.123	0.210	0.032	-0.050
	(0.053)	(0.104)	(0.132)	(0.116)	(0.163)	(0.200)	(0.148)	(0.258)
Average Score of Rater and SR (Scale: 2-8)	0.182***	0.362***	0.029	0.224***	0.285***	0.119	0.226***	0.195
	(0.031)	(0.047)	(0.060)	(0.043)	(0.074)	(0.099)	(0.080)	(0.128)
Competitive Category: Operations = 1								
Competitive Category: Operations Support = 1								
Competitive Category: Force Sustainment = 1								
Competitive Category: Army Medical Department = 1								
Year Promoted to Major	-0.063***	0.001	-0.015	-0.012	-0.042**	-0.152***	-0.024	-0.042
	(0.012)	(0.011)	(0.013)	(0.035)	(0.021)	(0.047)	(0.037)	(0.027)
Observations	529	330	90	146	397	164	219	225
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy	Pre-Post Dummy

Figure B.4: Sensitivity Check Results for full panel, excluding promotions to lieutenant colonel in 2021, Administrative, Human Resources, Logistics Occupational Groups and Nurses

Appendix C

SUPPORTING DOCUMENTATION FOR CHAPTER 3

Sample

The sample for this study is active duty Army Officers from 2014 through 2024. The sample is drawn from the Defense Manpower Database Center (DMDC) personnel records, which includes information on all active-duty Army officers. The sample is limited to officers who have been in the Army for at least three years and who have completed their initial training. While each officer might not have a subordinate, all officers in the dataset have two documented supervisors on their performance appraisals, a first line supervisor (1LS) "boss" and a second line supervisor (2LS), or a "bosses boss."This ensures that the sample is representative of the population of active-duty Army officers. The full panel used in this study is created by merging individual personnel records based on their unique individual identifier with their annual performance data, and behavioral health appointment records.

Data for this study is inclusive of data from January 2010 through May 2024, drawn from DOD databases within the Person-Event Data Environment (PDE). The Defense Manpower Data Center Active-Duty Master File provides quarterly snapshots of personal and professional information at the individual level, including a deidentified unique identifier (PID), demographics, and

career details. Performance data comes from the Army's Evaluation Entry System, which contain the employee's PID, occupational information, supervisor rankings, and evaluation periods. The primary medical database, also linked by an employee's PID, comprises the employee's medical appointment records.

My method of identification of behavioral health seeking behavior is through the Defense Health Agency's (DHA's) cost accounting of appointments in their system of record. Using functional cost codes for outpatient mental health clinic or health care appointments (see Table C.2), I identify whether employees have ever sought outpatient mental health support. I identify variations of these health seekers as:

- Never seeker — No inpatient or outpatient mental health, behavioral health, or psychological appointments of record.
- Only behavioral health seeker — Only seeks outpatient mental health appointments without seeking from any other FCC code defined in Table C.2.
- Any seekers — Seeks additional supportive care (e.g. in or out-patient psychiatric care, substance abuse programs, support groups), as well as possible seeking of mental health appointments.

The selection of the sample is based on employees that meet the following criteria. First, they must be behavioral health seekers only, or never seekers. Any seekers, who are or have pursued, additional wellness seeking behaviors are beyond the scope of this paper and are removed from the sample. That is at the specified time of observation in their military careers, they have either only sought outpatient behavioral health appointments as identified by their functional cost code (see Table C.2). Additionally, the employees must have a military officer (not a Department

of Defense Civilian) as a first and second-line supervisor to remain in the sample. Through this selection process, 4,456 officers who are in their third year of service serve as the primary sample for this study.

List of Variables, Functional Form, and Data Source

Table C.1 lists the variables used for regressions in this paper, lists their quantitative functional form, and an explanation of how the variable was coded and the data source for the variable. The variables are grouped into three categories: dependent variable, key explanatory variables, and controls. The dependent variable is the standardized performance score of the employee while at the specified years of service. The key explanatory variables are the behavioral health seeking indicators for the employee and their first and second line supervisors over a specified period of previous years. The control variables include demographic information about the employee, as well as characteristics of the supervisor relationship.

Research Design - Using Medical Expense and Performance Reporting System Codes

The DHA uses Medical Expense and Performance Reporting System (MEPRS) codes (also known as functional cost codes) to account for all workload, expense, and full-time equivalent data (Defense Health Agency, 2018). This study utilizes the three-digit MEP codes ”BFD” and ”BFZ,” representing Outpatient Mental Health Clinic visits, as the identification strategy for an employee’s mental or behavioral health actions. Functional Cost Codes (FCC) are used to classify the type of outpatient mental health services provided in the military healthcare system. All other codes were used to additionally classify ”any seekers.” The following table lists the relevant FCCs for this study:

Table C.1: List of Variables, Functional Form, and Data Source

Variable	Functional Form	Source; Detail
Dependent Variable		
Average Performance Score (Standardized)	Binary	Evaluation Entry System or Officer Evaluation Record System; Officers average senior rater score while at the rank of major; mean of 0, standard deviation of 1
Key Explanatory Variables		
First and second-line supervisors seeking behavioral health in the last three years	Binary	Defense Health Agency historical appointment database; filtered data to Army officers, and looked for appointments that matched the corresponding functional cost code for Behavioral Health. Then matched to unique ID of first and second line
Controls		
Not White	Binary	DMDC Master File; Binary indicator of race (1 indicates not-white; 0 indicates white)
Not Female	Binary	DMDC Master File; Binary indicator of gender (1 indicates not-female; 0 indicates male)
Age	Continuous	DMDC Master File; Years of Age when Years of Service = 3
Individuals Not Married	Binary	DMDC Master File; Individual Married when eligible for promotion to lieutenant colonel (0 indicates married; 1 indicates never married, divorced, or widowed)
Employee has only sought BH Health	Binary	Defense Health Agency historical appointment database; filtered data to Army officers, and looked for appointments that matched the corresponding functional cost code for outpatient behavioral health appointments.
Employee has sought BH Health in last 3 years	Binary	Defense Health Agency historical appointment database; filtered data to Army officers, and looked for appointments that matched the corresponding functional cost code for outpatient behavioral health appointments, within the last three years.
Length of supervisor relationship	Standardized	Number of months employee has been supervised by current supervisor.
Employee and Supervisor are same gender	Binary	DMDC Master File; Employee and supervisor are the same gender, male or female (Yes = 1, No = 0).
Employee and Supervisor are same race	Binary	DMDC Master File; Employee and supervisor are the same race, white or non-white (Yes = 1, No = 0).
Employee and Supervisor are same race and gender	Binary	DMDC Master File; Employee and supervisor are the same gender, male or female, and race, white or non-white (Yes = 1, No = 0).
Employee and 2LS are same gender	Binary	DMDC Master File; Employee and supervisor are the same gender, male or female (Yes = 1, No = 0).
Employee and 2LS are same race	Binary	DMDC Master File; Employee and supervisor are the same race, white or non-white (Yes = 1, No = 0).
Employee and 2LS are same race and gender	Binary	DMDC Master File; Employee and supervisor are the same gender, male or female, and race, white or non-white (Yes = 1, No = 0).
First line supervisor span of control	Standardized	Evaluation Entry System or Officer Evaluation Record System; After matching supervisors to employees, counting the number of times within one year that an individual is on a performance record as a supervisor.
Second line supervisor span of control	Standardized	Evaluation Entry System or Officer Evaluation Record System; After matching supervisors to employees, counting the number of times within one year that an individual is on a performance record as a second-line supervisor.

Table C.2: Medical Expense and Performance Reporting System (MEPRS) Functional Cost Codes

Functional Cost Codes (FCC)	**Title**
BFA_	Outpatient Psychiatry Clinic
BFB_	Outpatient Psychology Clinic
BFC_	Outpatient Child Guidance Clinic
BFD_	Outpatient Mental Health Clinic
BFE_	Outpatient Social Work Clinic
BFF_	Outpatient Substance Abuse Clinic
BFX_	Outpatient Psychiatric/Mental Healthcare Cost Pools
BFZ_	Outpatient Psychiatric/Mental Healthcare Not Elsewhere Classified (Requires Advanced Approval)

Note: MEPRS functional cost codes can be found in Appendix 5 — Table 5 of (Defense Health Agency, 2018).

Appendix D

ADDITIONAL TABLES AND FIGURES FOR CHAPTER 3

Expanded results of the primary analysis for 7 years of service

The descriptive statistics for employees with seven years of service can be found in Table D.1 At seven years of service the sample size has decreased significantly to 1,519 employees. This can be attributed to it being a smaller sample of employees who have been in the system for 7 years that have only used behavioral health services or no services at all. In the sample, 938 employees have supervisors who have not had behavioral health appointments in the last three years, 93 employees had both supervisors seek BH in the last three years, and 488 employees had one supervisor seek BH in the last three years.

Sensitivity Check for Seven Years of Service

The results show that the model is robust to this specification, with similar magnitude and direction to the variables in the primary model, at seven years of service. Model (1) in Table D.12 shows that an employee's probability of getting at least one most qualified rating decreases 7% if both supervisors have sought BH in the last three years (b=-0.072, $p < 0.05$). Similar to the sensitivity check at three years of service, Table D.12 shows that there is a negative average marginal

Table D.1: Descriptive statistics for employees with 7 years of service

Descriptive Statistics by Group											
	Full Sample										
	Count	Mean	SD	Min	Max	Mean (Neither)	SD (Neither)	Mean (Both)	SD (Both)	Mean (One)	SD (One)
Composite Eval Score by Year Standardized (Z-Score)	1519	0.073	1.090	-5.198	1.226	0.093	1.115	-0.034	1.067	0.054	1.043
Received at least one 'most qualified' rating from 2LS	1519	0.694	0.461	0.000	1.000	0.708	0.455	0.613	0.490	0.682	0.466
All not white individuals (=1)	1519	0.257	0.437	0.000	1.000	0.258	0.438	0.258	0.440	0.256	0.437
All females (=1)	1519	0.134	0.341	0.000	1.000	0.138	0.345	0.151	0.360	0.125	0.331
Age (Years)	1519	31.426	3.385	25.000	51.000	31.268	3.196	31.892	3.279	31.641	3.727
Individuals not married (=1)	1519	0.326	0.469	0.000	1.000	0.332	0.471	0.194	0.397	0.340	0.474
Has only sought BH help; no other categories	1519	0.425	0.495	0.000	1.000	0.405	0.491	0.559	0.499	0.439	0.497
Individual had BH appt in last 3 years	1519	0.250	0.433	0.000	1.000	0.235	0.424	0.323	0.470	0.266	0.443
Individual had more than the mean BH appts (year)	1519	0.022	0.146	0.000	1.000	0.019	0.137	0.043	0.204	0.023	0.149
Rank Grouping	1519	1.095	0.293	1.000	2.000	1.088	0.284	1.129	0.337	1.100	0.301
Length of supervisor relationships (months)	1519	12.003	5.546	2.000	44.000	12.101	5.377	12.072	5.492	11.800	5.872
Employee and supervisor are same gender (binary)	1519	0.779	0.415	0.000	1.000	0.790	0.408	0.720	0.451	0.768	0.422
Employee and supervisor are same race (binary)	1519	0.601	0.490	0.000	1.000	0.597	0.491	0.473	0.502	0.633	0.482
Employee and supervisor match race and gender	1519	0.489	0.500	0.000	1.000	0.500	0.500	0.344	0.478	0.496	0.500
Employee and 2LS are same gender (binary)	1519	0.788	0.409	0.000	1.000	0.793	0.405	0.742	0.440	0.787	0.410
Employee and 2LS are same race (binary)	1519	0.652	0.477	0.000	1.000	0.648	0.478	0.645	0.481	0.660	0.474
Employee and 2LS match race and gender (binary)	1519	0.530	0.499	0.000	1.000	0.533	0.499	0.516	0.502	0.527	0.500
1st Line Supervisor Span of Control (count)	1519	5.402	3.354	1.000	37.000	5.430	3.198	5.419	3.453	5.344	3.623
2nd Line Supervisor Span of Control (count)	1519	32.748	23.949	1.000	101.000	32.967	23.733	27.129	22.426	33.398	24.547
						938		93		488	

Note: The table presents descriptive statistics for the full sample (N=1,159). The group means are identified by if their first or second line supervisor has sought behavioral health in the last three years. First and second line supervisor behavioral health seeking behavior is indicated by: neither supervisor has, one of the supervisors has, or both of them have. (Neither, Both, and One). SD = Standard Deviation. BH = Behavioral Health. 2LS = Second Line Supervisor.

effect for employees that have more than the mean BH appointments per year, with the greatest effect occurring at the intensive margins. Unlike the three-year model in Table D.11 Model (5), in Table D.12 Model (5), the first line supervisors BH seeking intensity has an inconclusive effect on employee performance. This suggests that the first line supervisor's BH seeking behavior having a negative effect on the employee's performance rating may be conditional on the average first line supervisors years of service and rank.

Table D.2: Primary regression results for employees with 7 years of service

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 7	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 7	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 7	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 7	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 7
IV: Extensive Margin Indicator (BH in 3 years) = 1	-0.056			0.741**	
	(0.116)			(0.350)	
IV: Robust Intensive Margin Indicator (BH in 3 years) = 1		-0.060			
		(0.116)			
IV: Robust Intensive Margin Indicator (BH in 3 years) = 2		-0.012			0.085
		(0.059)			(0.124)
IV: Narrow Intensive Margin Indicator (BH in 3 years) = 1			-0.087		
			(0.124)		
1LS had more than mean BH appointments = 1					0.105
					(0.152)
Individual had BH appt in last 3 years = 1	-0.208**	-0.210***	-0.213*	-0.204**	-0.212*
	(0.094)	(0.076)	(0.115)	(0.099)	(0.115)
Individual had more than the mean BH appts (year) = 1	-0.250	-0.427*	-0.541**	-0.270	-0.549**
	(0.303)	(0.232)	(0.263)	(0.341)	(0.261)
All not white individuals (=1) = 1	-0.097	-0.095	-0.219*	-0.030	-0.217*
	(0.097)	(0.077)	(0.118)	(0.098)	(0.118)
All females (=1) = 1	0.164	0.068	-0.033	0.167	-0.032
	(0.112)	(0.093)	(0.153)	(0.118)	(0.153)
Individuals not married (=1) = 1	-0.121	-0.168***	-0.254***	-0.123	-0.252***
	(0.076)	(0.061)	(0.094)	(0.079)	(0.095)
Rank Grouping = 2, 2	-0.478***	-0.496***	-0.444***	-0.486***	-0.445***
	(0.131)	(0.105)	(0.165)	(0.137)	(0.165)
Employee and supervisor are same gender (binary) = 1	0.006	-0.058	-0.098	-0.009	-0.103
	(0.134)	(0.104)	(0.138)	(0.148)	(0.137)
Employee and supervisor are same race (binary) = 1	0.030	-0.068	-0.238	0.052	-0.239
	(0.158)	(0.116)	(0.147)	(0.171)	(0.147)
Employee and supervisor match race and gender = 1	0.169	0.139	0.016	0.194	0.019
	(0.178)	(0.134)	(0.181)	(0.192)	(0.181)
Span of control (1LS), Standardized	0.002	0.001	0.002	0.002	0.003
	(0.026)	(0.019)	(0.023)	(0.028)	(0.023)
Span of control (2LS), Standardized	0.017	0.030**	0.032	0.027	0.031
	(0.018)	(0.014)	(0.023)	(0.018)	(0.023)
Occupational workgroup = 2, Operations Support	0.107	0.098	0.425	-0.024	0.428
	(0.246)	(0.205)	(0.331)	(0.251)	(0.331)
Occupational workgroup = 3, Force Sustainment	0.070	-0.006	0.057	-0.026	0.052
	(0.130)	(0.104)	(0.164)	(0.136)	(0.165)
Occupational workgroup = 4, Medical	-0.165	0.044	0.409*	-0.196	0.410*
	(0.187)	(0.142)	(0.209)	(0.189)	(0.209)
Occupational workgroup = 5, Other	0.031	0.042	0.235	-0.049	0.233
	(0.115)	(0.092)	(0.148)	(0.117)	(0.149)
Total supervisor relationship length, Standardized	0.195***	0.150***	0.147**	0.159***	0.148**
	(0.058)	(0.044)	(0.061)	(0.061)	(0.061)
Presidential Administration = 2, Trump	0.101	0.161*	0.173	0.104	0.166
	(0.121)	(0.092)	(0.120)	(0.131)	(0.121)
Presidential Administration = 3, Biden	0.159	0.206**	0.198	0.175	0.198
	(0.122)	(0.092)	(0.125)	(0.132)	(0.126)
Constant	-0.206	-0.106	-0.021	-0.178	-0.109
	(0.209)	(0.167)	(0.225)	(0.229)	(0.232)
Observations	1,031	1,519	581	944	581
R-squared	0.057	0.055	0.102	0.055	0.102
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents primary regression results for employees with 3 years of service. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Expanded results of the primary analysis for 11 years of service

The descriptive statistics for employees with seven years of service can be found in Table D.3 At eleven years of service the sample size has further reduced to 265 employees. This can be attributed to it being a smaller sample of employees who have been in the system for 7 years that have only used behavioral health services or no services at all. In the sample, 168 employees have supervisors who have not had behavioral health appointments in the last three years, 16 employees had both supervisors seek BH in the last three years, and 81 employees had one supervisor seek BH in the last three years.

Table D.3: Descriptive statistics for employees with 11 years of service

Descriptive Statistics by Group											
	Full Sample										
Count	Mean	SD	Min	Max	Mean (Neither)	SD (Neither)	Mean (Both)	SD (Both)	Mean (One)	SD (One)	
Composite Eval Score by Year Standardized (Z-Score)	265	-0.071	0.949	-3.057	1.226	-0.048	0.955	-0.213	0.846	-0.089	0.962
Received at least one 'most qualified' rating from 2LS	265	0.623	0.486	0.000	1.000	0.613	0.488	0.750	0.447	0.617	0.489
All not white individuals (=1)	265	0.343	0.476	0.000	1.000	0.351	0.479	0.375	0.500	0.321	0.470
All females (=1)	265	0.087	0.282	0.000	1.000	0.071	0.258	0.063	0.250	0.123	0.331
Age (Years)	265	36.264	3.825	26.000	52.000	36.268	3.712	35.563	2.683	36.395	4.248
Individuals not married (=1)	265	0.121	0.326	0.000	1.000	0.125	0.332	0.000	0.000	0.136	0.345
Has only sought BH help; no other categories	265	0.498	0.501	0.000	1.000	0.518	0.501	0.438	0.512	0.469	0.502
Individual had BH appt in last 3 years	265	0.268	0.444	0.000	1.000	0.256	0.438	0.313	0.479	0.284	0.454
Individual had more than the mean BH appts (year)	265	0.038	0.191	0.000	1.000	0.036	0.186	0.063	0.250	0.037	0.190
Rank Grouping	265	1.430	0.496	1.000	2.000	1.440	0.498	1.438	0.512	1.407	0.494
Length of supervisor relationships (months)	265	12.889	7.227	4.000	59.000	12.994	7.063	12.594	6.301	12.728	7.792
Employee and supervisor are same gender (binary)	265	0.766	0.424	0.000	1.000	0.792	0.407	0.875	0.342	0.691	0.465
Employee and supervisor are same race (binary)	265	0.558	0.498	0.000	1.000	0.560	0.498	0.688	0.479	0.531	0.502
Employee and supervisor match race and gender	265	0.445	0.498	0.000	1.000	0.458	0.500	0.625	0.500	0.383	0.489
Employee and 2LS are same gender (binary)	265	0.842	0.366	0.000	1.000	0.887	0.318	0.813	0.403	0.753	0.434
Employee and 2LS are same race (binary)	265	0.634	0.483	0.000	1.000	0.577	0.495	0.688	0.479	0.741	0.441
Employee and 2LS match race and gender (binary)	265	0.528	0.500	0.000	1.000	0.524	0.501	0.500	0.516	0.543	0.501
1st Line Supervisor Span of Control (count)	265	5.087	4.397	1.000	37.000	4.589	3.468	6.500	7.099	5.840	5.269
2nd Line Supervisor Span of Control (count)	265	26.721	21.663	1.000	89.000	26.196	20.882	26.875	27.461	27.778	22.255
						168		16		81	

Note: The table presents descriptive statistics for the full sample (N=265). The group means are identified by if their first or second line supervisor has sought behavioral health in the last three years. First and second line supervisor behavioral health seeking behavior is indicated by: neither supervisor has, one of the supervisors has, or both of them have. (Neither, Both, and One). SD = Standard Deviation. BH = Behavioral Health. 2LS = Second Line Supervisor.

Table D.4: Primary regression results for employees with 11 years of service

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 11	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 11	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 11	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 11	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 11
IV:Extensive Margin Indicator (BH in 3 years) = 1	-0.131				
	(0.253)				
IV: Robust Intensive Margin Indicator (BH in 3 years) = 1		-0.088			
		(0.253)			
IV: Robust Intensive Margin Indicator (BH in 3 years) = 2		-0.067			0.158
		(0.127)			(0.279)
IV: Narrow Intensive Margin Indicator (BH in 3 years) = 1			-0.146		
			(0.280)		
IV: Extensive Margin Indicator (BH in 3 years) = 0, omitted				-	
1LS had more than mean BH appointments = 1					-0.706*
					(0.416)
Individual had BH appt in last 3 years = 1	-0.536***	-0.319**	0.203	-0.609***	0.201
	(0.183)	(0.147)	(0.248)	(0.194)	(0.252)
Individual had more than the mean BH appts (year) = 1	0.212	0.084	0.213	0.276	0.392
	(0.397)	(0.293)	(0.360)	(0.437)	(0.401)
All not white individuals (=1) = 1	-0.170	-0.257**	-0.223	-0.270	-0.202
	(0.165)	(0.128)	(0.228)	(0.172)	(0.232)
All females (=1) = 1	0.383*	0.467***	0.421	0.351	0.352
	(0.223)	(0.174)	(0.358)	(0.230)	(0.372)
Individuals not married (=1) = 1	0.176	-0.130	-0.889***	0.178	-0.843***
	(0.211)	(0.165)	(0.275)	(0.220)	(0.278)
Rank Grouping = 2, 2	-0.318**	-0.259**	-0.277	-0.302*	-0.305
	(0.154)	(0.131)	(0.245)	(0.162)	(0.253)
Employee and supervisor are same gender (binary) = 1	0.249	0.166	0.068	0.208	-0.009
	(0.217)	(0.168)	(0.292)	(0.226)	(0.302)
Employee and supervisor are same race (binary) = 1	0.373	0.243	0.251	0.335	0.181
	(0.276)	(0.232)	(0.408)	(0.291)	(0.420)
Employee and supervisor match race and gender = 1	-0.448	-0.328	-0.196	-0.523	-0.120
	(0.345)	(0.279)	(0.465)	(0.360)	(0.474)
Span of control (1LS), Standardized	0.103**	0.012	-0.007	0.113**	-0.010
	(0.042)	(0.035)	(0.039)	(0.053)	(0.039)
Span of control (2LS), Standardized	-0.032	-0.015	-0.030	-0.031	-0.026
	(0.040)	(0.031)	(0.053)	(0.041)	(0.054)
Occupational workgroup = 2, Operations Support	-0.669	-0.349	0.041	-0.603	0.006
	(0.412)	(0.362)	(0.640)	(0.381)	(0.650)
Occupational workgroup = 3, Force Sustainment	-0.258	-0.318	-0.543	-0.150	-0.577
	(0.225)	(0.201)	(0.383)	(0.239)	(0.388)
Occupational workgroup = 4, Medical	-0.327	-0.120	-0.132	-0.122	-0.196
	(0.397)	(0.279)	(0.404)	(0.410)	(0.425)
Occupational workgroup = 5, Other	-0.343*	-0.323**	-0.329	-0.296	-0.350
	(0.181)	(0.160)	(0.309)	(0.184)	(0.312)
Total supervisor relationship length, Standardized	0.058	0.095	0.196*	0.061	0.179
	(0.087)	(0.069)	(0.104)	(0.090)	(0.108)
Presidential Administration = 2, Trump	0.120	0.147	0.265	0.225	0.214
	(0.207)	(0.167)	(0.281)	(0.213)	(0.290)
Presidential Administration = 3, Biden	0.219	0.198	0.113	0.339	0.076
	(0.195)	(0.156)	(0.276)	(0.208)	(0.285)
Constant	-0.009	0.133	0.073	-0.044	0.061
	(0.328)	(0.263)	(0.438)	(0.380)	(0.604)
Observations	184	265	97	168	97
R-squared	0.147	0.097	0.190	0.175	0.198
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents primary regression results for employees with 11 years of service. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Temporal Extension for Behavioral Health Seeking in the Last Two Years

This section includes regression results evaluating how employee and supervisor health seeking effects results if it has been in the last two years instead of the last three years. Results are reported for 3, 7, and 11 years of service.

To further check the heterogeneity of the sample, I evaluate combinations of employees and supervisors that have had behavioral health appointments in the last two years, with employees at three, seven, and eleven years of service. For all three years of service groups, I find that the results are similar to their primary models in magnitude and direction (See Table D.5, Table D.6, and Table D.7, respectively for results).

Table D.5: Regression results for employees with 3 years of service; BH Seeking in last two years

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 3	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 3	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 3	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 3	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 3
IV: Robust Intensive Margin Indicator (BH in 2 years) = 1		-0.076			
		(0.072)			
IV: Robust Intensive Margin Indicator (BH in 2 years) = 2		-0.084**			0.005
		(0.035)			(0.076)
IV: Extensive Margin Indicator (BH in 2 years) = 1	-0.058			-0.197	
	(0.073)			(0.187)	
IV: Narrow Intensive Margin Indicator (BH in 2 years) = 1			-0.009		
			(0.076)		
Individual had BH appt in last 3 years = 1	-0.394***	-0.320***	-0.203**	-0.406***	-0.204***
	(0.064)	(0.052)	(0.079)	(0.069)	(0.079)
Indvidual had more than the mean BH appts (year) = 1	-0.222*	-0.243**	-0.178	-0.235*	-0.177
	(0.122)	(0.109)	(0.209)	(0.127)	(0.209)
All not white individuals (=1) = 1	-0.210***	-0.217***	-0.256***	-0.207***	-0.254***
	(0.044)	(0.037)	(0.066)	(0.045)	(0.066)
All females (=1) = 1	0.094	0.088*	0.079	0.103*	0.077
	(0.058)	(0.047)	(0.074)	(0.061)	(0.075)
Individuals not married (=1) = 1	-0.043	-0.062**	-0.109**	-0.048	-0.107*
	(0.037)	(0.031)	(0.055)	(0.038)	(0.055)
Rank Grouping = 2, 2	0.015	-0.194	-0.833**	0.156	-0.839**
	(0.384)	(0.314)	(0.350)	(0.405)	(0.352)
Employee and supervisor are same gender (binary) = 1	-0.026	-0.032	-0.013	-0.026	-0.012
	(0.067)	(0.056)	(0.092)	(0.070)	(0.092)
Employee and supervisor are same race (binary) = 1	-0.022	-0.024	0.005	-0.036	0.008
	(0.072)	(0.059)	(0.095)	(0.075)	(0.095)
Employee and supervisor match race and gender = 1	0.053	0.042	-0.050	0.074	-0.053
	(0.082)	(0.069)	(0.114)	(0.086)	(0.114)
Span of control (1LS), Standardized	-0.045***	-0.045***	-0.051***	-0.043***	-0.052***
	(0.014)	(0.011)	(0.016)	(0.015)	(0.016)
Span of control (2LS), Standardized	0.032**	0.030**	0.025	0.035**	0.024
	(0.014)	(0.012)	(0.021)	(0.014)	(0.021)
Occupational workgroup = 2, Operations Support	-0.156	-0.141	-0.216	-0.143	-0.218
	(0.162)	(0.133)	(0.212)	(0.176)	(0.212)
Occupational workgroup = 3, Force Sustainment	0.091	0.092	0.000	0.119	0.001
	(0.075)	(0.062)	(0.103)	(0.078)	(0.103)
Occupational workgroup = 4, Medical	0.057	0.040	-0.031	0.058	-0.030
	(0.084)	(0.069)	(0.114)	(0.087)	(0.114)
Occupational workgroup = 5, Other	0.166**	0.133**	0.027	0.168**	0.026
	(0.065)	(0.054)	(0.092)	(0.067)	(0.092)
Total supervisor relationship length, Standardized	0.133***	0.131***	0.122***	0.131***	0.124***
	(0.027)	(0.023)	(0.043)	(0.028)	(0.043)
Presidential Administration = 2, Trump	-0.114**	-0.101**	-0.099	-0.107**	-0.095
	(0.048)	(0.040)	(0.071)	(0.049)	(0.071)
Presidential Administration = 3, Biden	-0.063	-0.038	-0.023	-0.052	-0.024
	(0.049)	(0.041)	(0.070)	(0.052)	(0.070)
1LS had more than mean BH appointments = 1					-0.084
					(0.091)
Constant	-0.002	0.030	0.085	-0.022	0.086
	(0.104)	(0.086)	(0.137)	(0.108)	(0.150)
Observations	3,247	4,456	1,373	3,083	1,373
R-squared	0.049	0.046	0.041	0.049	0.042
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with three years of service, and employees and supervisors who have sought BH in the last two years. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Table D.6: Regression results for employees with 7 years of service; BH Seeking in last two years

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 7	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 7	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 7	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 7	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 7
IV: Robust Intensive Margin Indicator (BH in 2 years) = 1		0.064			
		(0.119)			
IV: Robust Intensive Margin Indicator (BH in 2 years) = 2		0.007			-0.034
		(0.063)			(0.130)
IV: Extensive Margin Indicator (BH in 2 years) = 1	0.065			0.736**	
	(0.119)			(0.331)	
IV: Narrow Intensive Margin Indicator (BH in 2 years) = 1			0.034		
			(0.131)		
Individual had BH appt in last 3 years = 1	-0.223**	-0.213***	-0.187	-0.218**	-0.183
	(0.089)	(0.076)	(0.128)	(0.094)	(0.127)
Indvidual had more than the mean BH appts (year) = 1	-0.210	-0.432*	-0.598**	-0.240	-0.604**
	(0.312)	(0.231)	(0.270)	(0.339)	(0.268)
All not white individuals (=1) = 1	-0.093	-0.093	-0.175	-0.059	-0.171
	(0.090)	(0.077)	(0.130)	(0.093)	(0.130)
All females (=1) = 1	0.162	0.069	-0.078	0.150	-0.076
	(0.107)	(0.093)	(0.172)	(0.111)	(0.172)
Individuals not married (=1) = 1	-0.167**	-0.165***	-0.167	-0.165**	-0.164
	(0.071)	(0.061)	(0.110)	(0.074)	(0.111)
Rank Grouping = 2, 2	-0.570***	-0.500***	-0.312	-0.546***	-0.313
	(0.124)	(0.105)	(0.193)	(0.127)	(0.194)
Employee and supervisor are same gender (binary) = 1	0.028	-0.058	-0.168	0.017	-0.176
	(0.128)	(0.104)	(0.158)	(0.137)	(0.158)
Employee and supervisor are same race (binary) = 1	0.108	-0.069	-0.385**	0.097	-0.387**
	(0.149)	(0.116)	(0.163)	(0.158)	(0.163)
Employee and supervisor match race and gender = 1	0.094	0.144	0.086	0.117	0.089
	(0.168)	(0.134)	(0.205)	(0.177)	(0.204)
Span of control (1LS), Standardized	-0.000	0.000	0.003	0.001	0.005
	(0.024)	(0.019)	(0.026)	(0.026)	(0.027)
Span of control (2LS), Standardized	0.025	0.031**	0.035	0.029*	0.034
	(0.016)	(0.014)	(0.026)	(0.017)	(0.026)
Occupational workgroup = 2, Operations Support	0.017	0.094	0.555	-0.055	0.560
	(0.238)	(0.204)	(0.373)	(0.242)	(0.373)
Occupational workgroup = 3, Force Sustainment	-0.031	-0.015	0.116	-0.076	0.110
	(0.122)	(0.104)	(0.189)	(0.126)	(0.190)
Occupational workgroup = 4, Medical	-0.090	0.039	0.452*	-0.121	0.453*
	(0.162)	(0.141)	(0.254)	(0.167)	(0.254)
Occupational workgroup = 5, Other	0.003	0.038	0.220	-0.030	0.219
	(0.107)	(0.092)	(0.169)	(0.108)	(0.169)
Total supervisor relationship length, Standardized	0.170***	0.150***	0.158**	0.148***	0.160**
	(0.053)	(0.044)	(0.071)	(0.055)	(0.071)
Presidential Administration = 2, Trump	0.127	0.163*	0.151	0.125	0.139
	(0.112)	(0.092)	(0.138)	(0.118)	(0.139)
Presidential Administration = 3, Biden	0.199*	0.208**	0.160	0.197	0.160
	(0.113)	(0.092)	(0.141)	(0.120)	(0.141)
1LS had more than mean BH appointments = 1					0.109
					(0.157)
Constant	-0.219	-0.119	0.030	-0.188	0.060
	(0.197)	(0.165)	(0.244)	(0.212)	(0.270)
Observations	1,132	1,519	447	1,071	447
R-squared	0.061	0.055	0.102	0.058	0.103
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with seven years of service, and employees and supervisors who have sought BH in the last two years. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Table D.7: Regression results for employees with 11 years of service; BH Seeking in last two years

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 11	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 11	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 11	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 11	(5) OLS Eval Score (YR STD) Int. Margin (Robust - High Intensity) Employee YOS = 11
IV: Robust Intensive Margin Indicator (BH in 2 years) = 1		0.022			
		(0.288)			
IV: Robust Intensive Margin Indicator (BH in 2 years) = 2		-0.079			0.049
		(0.141)			(0.317)
IV: Extensive Margin Indicator (BH in 2 years) = 1	-0.054				
	(0.311)				
IV: Narrow Intensive Margin Indicator (BH in 2 years) = 1			-0.035		
			(0.311)		
IV: Extensive Margin Indicator (BH in 2 years) = 0, omitted				-	
Individual had BH appt in last 3 years = 1	-0.533***	-0.316**	0.422	-0.608***	0.420
	(0.168)	(0.147)	(0.275)	(0.176)	(0.281)
Indvidual had more than the mean BH appts (year) = 1	0.171	0.081	-0.091	0.213	0.095
	(0.386)	(0.292)	(0.426)	(0.432)	(0.473)
All not white individuals (=1) = 1	-0.251*	-0.262**	-0.125	-0.303*	-0.105
	(0.149)	(0.129)	(0.280)	(0.156)	(0.284)
All females (=1) = 1	0.451**	0.464***	0.311	0.428**	0.222
	(0.195)	(0.172)	(0.412)	(0.199)	(0.418)
Individuals not married (=1) = 1	0.122	-0.123	-1.017***	0.117	-0.969***
	(0.204)	(0.165)	(0.343)	(0.213)	(0.341)
Rank Grouping = 2, 2	-0.336**	-0.259**	-0.277	-0.362**	-0.304
	(0.149)	(0.131)	(0.279)	(0.156)	(0.288)
Employee and supervisor are same gender (binary) = 1	0.127	0.171	0.149	0.077	0.055
	(0.199)	(0.168)	(0.361)	(0.200)	(0.355)
Employee and supervisor are same race (binary) = 1	0.252	0.254	0.409	0.199	0.330
	(0.278)	(0.230)	(0.421)	(0.280)	(0.429)
Employee and supervisor match race and gender = 1	-0.358	-0.344	-0.398	-0.357	-0.313
	(0.341)	(0.278)	(0.493)	(0.347)	(0.497)
Span of control (1LS), Standardized	0.096**	0.010	0.006	0.102**	0.004
	(0.040)	(0.035)	(0.041)	(0.047)	(0.041)
Span of control (2LS), Standardized	-0.033	-0.013	-0.016	-0.032	-0.011
	(0.038)	(0.031)	(0.061)	(0.040)	(0.062)
Occupational workgroup = 2, Operations Support	-0.970***	-0.351	0.382	-0.946**	0.363
	(0.370)	(0.366)	(0.689)	(0.366)	(0.692)
Occupational workgroup = 3, Force Sustainment	-0.181	-0.321	-0.708*	-0.111	-0.749*
	(0.219)	(0.200)	(0.420)	(0.228)	(0.425)
Occupational workgroup = 4, Medical	-0.298	-0.137	0.077	-0.190	-0.001
	(0.328)	(0.277)	(0.508)	(0.331)	(0.544)
Occupational workgroup = 5, Other	-0.361**	-0.328**	-0.305	-0.328*	-0.323
	(0.171)	(0.159)	(0.352)	(0.172)	(0.352)
Total supervisor relationship length, Standardized	0.090	0.093	0.230**	0.086	0.215*
	(0.080)	(0.069)	(0.111)	(0.082)	(0.115)
Presidential Administration = 2, Trump	0.155	0.141	0.082	0.206	0.032
	(0.189)	(0.167)	(0.327)	(0.194)	(0.335)
Presidential Administration = 3, Biden	0.268	0.198	-0.074	0.323*	-0.122
	(0.181)	(0.156)	(0.319)	(0.190)	(0.329)
1LS had more than mean BH appointments = 1					-0.719
					(0.456)
Constant	0.098	0.132	-0.014	0.131	0.103
	(0.313)	(0.261)	(0.480)	(0.353)	(0.636)
Observations	198	265	78	186	78
R-squared	0.167	0.097	0.213	0.192	0.223
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with 11 years of service, and employees and supervisors who have sought BH in the last two years. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Temporal Extension for Behavioral Health Seeking in the Last Five Years

This section includes regression results evaluating how employee and supervisor health seeking effects results if it has been in the last five years instead of the last three years. Results are reported for 3, 7, and 11 years of service.

Finally, I evaluate combinations of employees and supervisors that have had behavioral health appointments in the last five years, with employees at three, seven, and eleven years of service. For all three years of service groups, I find that the results are similar to their primary models in magnitude and direction (See Table D.8, Table D.9, and Table D.10, respectively for results).

Table D.8: Regression results for employees with 3 years of service; BH Seeking in last five years

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 3	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 3	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 3	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 3	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 3
IV: Robust Intensive Margin Indicator (BH in 5 years) = 1		-0.025			
		(0.049)			
IV: Robust Intensive Margin Indicator (BH in 5 years) = 2		-0.062*			-0.041
		(0.032)			(0.049)
IV: Extensive Margin Indicator (BH in 5 years) = 1	-0.020			-0.065	
	(0.049)			(0.066)	
IV: Narrow Intensive Margin Indicator (BH in 5 years) = 1			0.035		
			(0.049)		
Individual had BH appt in last 3 years = 1	-0.374***	-0.329***	-0.271***	-0.395***	-0.270***
	(0.067)	(0.052)	(0.068)	(0.075)	(0.068)
Indvidual had more than the mean BH appts (year) = 1	-0.258*	-0.232**	-0.132	-0.306*	-0.134
	(0.155)	(0.109)	(0.133)	(0.166)	(0.133)
All not white individuals (=1) = 1	-0.175***	-0.218***	-0.273***	-0.172***	-0.270***
	(0.048)	(0.037)	(0.051)	(0.051)	(0.051)
All females (=1) = 1	0.027	0.088*	0.116*	0.025	0.114*
	(0.063)	(0.047)	(0.060)	(0.068)	(0.060)
Individuals not married (=1) = 1	-0.020	-0.062**	-0.093**	-0.017	-0.091**
	(0.041)	(0.031)	(0.041)	(0.044)	(0.041)
Rank Grouping = 2, 2	-0.220	-0.201	-0.332	-0.062	-0.335
	(0.415)	(0.319)	(0.453)	(0.463)	(0.453)
Employee and supervisor are same gender (binary) = 1	-0.125	-0.032	0.030	-0.136	0.030
	(0.077)	(0.056)	(0.071)	(0.083)	(0.071)
Employee and supervisor are same race (binary) = 1	-0.058	-0.027	0.002	-0.075	0.006
	(0.080)	(0.059)	(0.076)	(0.087)	(0.076)
Employee and supervisor match race and gender = 1	0.136	0.044	-0.056	0.167*	-0.059
	(0.094)	(0.069)	(0.088)	(0.101)	(0.088)
Span of control (1LS), Standardized	-0.015	-0.046***	-0.067***	-0.015	-0.067***
	(0.015)	(0.011)	(0.014)	(0.016)	(0.014)
Span of control (2LS), Standardized	0.020	0.031***	0.034**	0.027*	0.034**
	(0.015)	(0.012)	(0.016)	(0.016)	(0.016)
Occupational workgroup = 2, Operations Support	-0.234	-0.144	-0.168	-0.228	-0.169
	(0.173)	(0.134)	(0.174)	(0.192)	(0.174)
Occupational workgroup = 3, Force Sustainment	0.087	0.088	0.084	0.104	0.085
	(0.083)	(0.062)	(0.081)	(0.088)	(0.081)
Occupational workgroup = 4, Medical	0.124	0.037	-0.006	0.117	-0.003
	(0.092)	(0.069)	(0.092)	(0.097)	(0.092)
Occupational workgroup = 5, Other	0.205***	0.133**	0.078	0.206***	0.078
	(0.072)	(0.054)	(0.072)	(0.075)	(0.072)
Total supervisor relationship length, Standardized	0.112***	0.132***	0.159***	0.102***	0.160***
	(0.029)	(0.023)	(0.034)	(0.031)	(0.034)
Presidential Administration = 2, Trump	-0.105*	-0.091**	-0.095*	-0.079	-0.094*
	(0.054)	(0.040)	(0.053)	(0.058)	(0.053)
Presidential Administration = 3, Biden	-0.067	-0.036	-0.016	-0.054	-0.018
	(0.056)	(0.041)	(0.056)	(0.060)	(0.056)
1LS had more than mean BH appointments = 1					-0.125
					(0.089)
Constant	0.003	0.031	0.016	-0.018	0.062
	(0.114)	(0.087)	(0.111)	(0.123)	(0.118)
Observations	2,475	4,456	2,463	2,252	2,463
R-squared	0.046	0.046	0.050	0.047	0.051
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with three years of service, and employees and supervisors who have sought BH in the last five years. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Table D.9: Regression results for employees with 7 years of service; BH Seeking in last five years

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 7	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 7	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 7	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 7	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 7
IV: Robust Intensive Margin Indicator (BH in 5 years) = 1		-0.125			
		(0.091)			
IV: Robust Intensive Margin Indicator (BH in 5 years) = 2		-0.090			0.051
		(0.060)			(0.092)
IV: Extensive Margin Indicator (BH in 5 years) = 1	-0.101			-0.051	
	(0.093)			(0.128)	
IV: Narrow Intensive Margin Indicator (BH in 5 years) = 1			-0.051		
			(0.092)		
Individual had BH appt in last 3 years = 1	-0.191**	-0.209***	-0.227**	-0.183*	-0.225**
	(0.096)	(0.076)	(0.105)	(0.102)	(0.105)
Indvidual had more than the mean BH appts (year) = 1	-0.120	-0.420*	-0.512*	-0.061	-0.518*
	(0.303)	(0.230)	(0.275)	(0.365)	(0.274)
All not white individuals (=1) = 1	-0.121	-0.093	-0.153	-0.054	-0.150
	(0.105)	(0.077)	(0.102)	(0.107)	(0.102)
All females (=1) = 1	0.167	0.059	-0.029	0.177	-0.028
	(0.109)	(0.093)	(0.142)	(0.116)	(0.142)
Individuals not married (=1) = 1	-0.123	-0.170***	-0.222**	-0.128	-0.220**
	(0.080)	(0.061)	(0.087)	(0.083)	(0.087)
Rank Grouping = 2, 2	-0.447***	-0.492***	-0.511***	-0.465***	-0.513***
	(0.132)	(0.105)	(0.152)	(0.137)	(0.152)
Employee and supervisor are same gender (binary) = 1	-0.102	-0.060	0.014	-0.125	0.010
	(0.128)	(0.105)	(0.147)	(0.141)	(0.147)
Employee and supervisor are same race (binary) = 1	0.065	-0.060	-0.195	0.107	-0.196
	(0.145)	(0.116)	(0.161)	(0.158)	(0.161)
Employee and supervisor match race and gender = 1	0.106	0.129	0.118	0.121	0.121
	(0.173)	(0.134)	(0.182)	(0.187)	(0.182)
Span of control (1LS), Standardized	-0.016	0.001	0.025	-0.019	0.026
	(0.027)	(0.019)	(0.024)	(0.029)	(0.024)
Span of control (2LS), Standardized	0.029	0.030**	0.020	0.040**	0.019
	(0.019)	(0.014)	(0.020)	(0.019)	(0.020)
Occupational workgroup = 2, Operations Support	0.078	0.116	0.371	-0.102	0.374
	(0.282)	(0.205)	(0.277)	(0.283)	(0.278)
Occupational workgroup = 3, Force Sustainment	0.060	-0.001	0.080	-0.042	0.074
	(0.133)	(0.104)	(0.148)	(0.140)	(0.148)
Occupational workgroup = 4, Medical	-0.190	0.049	0.386**	-0.233	0.385**
	(0.191)	(0.142)	(0.187)	(0.193)	(0.187)
Occupational workgroup = 5, Other	0.094	0.045	0.130	0.016	0.127
	(0.118)	(0.092)	(0.132)	(0.121)	(0.133)
Total supervisor relationship length, Standardized	0.198***	0.146***	0.173***	0.163**	0.175***
	(0.061)	(0.044)	(0.053)	(0.064)	(0.053)
Presidential Administration = 2, Trump	0.101	0.174*	0.193*	0.107	0.190*
	(0.127)	(0.092)	(0.109)	(0.140)	(0.109)
Presidential Administration = 3, Biden	0.192	0.202**	0.155	0.219	0.155
	(0.123)	(0.092)	(0.119)	(0.133)	(0.119)
1LS had more than mean BH appointments = 1					0.133
					(0.152)
Constant	-0.125	-0.067	-0.201	-0.104	-0.255
	(0.212)	(0.166)	(0.214)	(0.230)	(0.211)
Observations	897	1,519	786	810	786
R-squared	0.064	0.056	0.080	0.059	0.081
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with seven years of service, and employees and supervisors who have sought BH in the last five years. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Table D.10: Regression results for employees with 11 years of service; BH Seeking in last five years

VARIABLES	(1) OLS Eval Score (YR STD) Extensive Margin Employee YOS = 11	(2) OLS Eval Score (YR STD) Intensive Margin (Robust) Employee YOS = 11	(3) OLS Eval Score (YR STD) Intensive Margin Narrow Employee YOS = 11	(4) OLS Eval Score (YR STD) Extensive Margin (High Intensity) Employee YOS = 11	(5) OLS Eval Score (YR STD) Int. Margin (Robust — High Intensity) Employee YOS = 11
IV: Robust Intensive Margin Indicator (BH in 5 years) = 1		0.088			
		(0.198)			
IV: Robust Intensive Margin Indicator (BH in 5 years) = 2		-0.073			-0.133
		(0.129)			(0.209)
IV: Extensive Margin Indicator (BH in 5 years) = 1	0.025			0.214	
	(0.214)			(0.316)	
IV: Narrow Intensive Margin Indicator (BH in 5 years) = 1			0.158		
			(0.209)		
Individual had BH appt in last 3 years = 1	-0.601***	-0.316**	-0.051	-0.645***	-0.058
	(0.203)	(0.147)	(0.206)	(0.222)	(0.209)
Indvidual had more than the mean BH appts (year) = 1	-0.179	0.079	0.671	-0.141	0.878**
	(0.274)	(0.299)	(0.431)	(0.288)	(0.416)
All not white individuals (=1) = 1	-0.121	-0.260**	-0.265	-0.250	-0.249
	(0.186)	(0.128)	(0.171)	(0.195)	(0.174)
All females (=1) = 1	0.643**	0.489***	0.516**	0.653*	0.458*
	(0.308)	(0.182)	(0.244)	(0.353)	(0.247)
Individuals not married (=1) = 1	0.126	-0.129	-0.416*	0.094	-0.353
	(0.226)	(0.163)	(0.247)	(0.241)	(0.246)
Rank Grouping = 2, 2	-0.301*	-0.269**	-0.279	-0.270	-0.316*
	(0.173)	(0.131)	(0.184)	(0.184)	(0.188)
Employee and supervisor are same gender (binary) = 1	0.058	0.197	0.062	0.105	-0.033
	(0.247)	(0.171)	(0.254)	(0.265)	(0.258)
Employee and supervisor are same race (binary) = 1	0.222	0.266	0.122	0.255	0.032
	(0.308)	(0.228)	(0.353)	(0.335)	(0.360)
Employee and supervisor match race and gender = 1	-0.114	-0.368	-0.245	-0.312	-0.145
	(0.395)	(0.277)	(0.385)	(0.430)	(0.390)
Span of control (1LS), Standardized	0.107**	0.010	-0.036	0.118**	-0.037
	(0.045)	(0.035)	(0.035)	(0.059)	(0.035)
Span of control (2LS), Standardized	-0.028	-0.014	-0.012	-0.026	-0.006
	(0.043)	(0.031)	(0.044)	(0.044)	(0.044)
Occupational workgroup = 2, Operations Support	-0.922**	-0.350	0.232	-0.830**	0.188
	(0.370)	(0.361)	(0.606)	(0.354)	(0.613)
Occupational workgroup = 3, Force Sustainment	-0.381	-0.336*	-0.268	-0.264	-0.319
	(0.238)	(0.203)	(0.289)	(0.250)	(0.290)
Occupational workgroup = 4, Medical	-0.271	-0.156	-0.247	-0.089	-0.304
	(0.408)	(0.284)	(0.356)	(0.427)	(0.360)
Occupational workgroup = 5, Other	-0.489**	-0.332**	-0.127	-0.413**	-0.141
	(0.195)	(0.159)	(0.240)	(0.206)	(0.234)
Total supervisor relationship length, Standardized	0.135*	0.097	0.038	0.147*	0.025
	(0.080)	(0.068)	(0.098)	(0.084)	(0.097)
Presidential Administration = 2, Trump	0.182	0.151	0.205	0.275	0.140
	(0.227)	(0.166)	(0.241)	(0.241)	(0.246)
Presidential Administration = 3, Biden	0.434**	0.206	0.004	0.574***	-0.047
	(0.188)	(0.155)	(0.245)	(0.206)	(0.249)
1LS had more than mean BH appointments = 1					-1.119***
					(0.356)
Constant	-0.027	0.121	0.202	-0.145	0.507
	(0.350)	(0.257)	(0.363)	(0.441)	(0.429)
Observations	152	265	140	136	140
R-squared	0.199	0.098	0.122	0.250	0.138
Robust	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with 11 years of service, and employees and supervisors who have sought BH in the last five years. The dependent variable is the standardized evaluation score (Z-Score) by year. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes robust standard errors. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Sensitivity check for operationalization of the dependent variable

In this check, I change the operationalization of the dependent variable from a standardized score to a binary variable. The binary variable is coded as 1 if the employee receives at least one 'most qualified' rating from their second line supervisor while at their current rank, and 0 otherwise. The number of 'most qualified' ratings an employee receives from their second line supervisor is a key determining factor in selection to promotion for higher ranks. This is a more stringent test of the hypothesis that behavioral health seeking is positively correlated with performance. Results are reported for 3, 7, and 11 years of service.

Table D.11: Regression results for employees with 3 years of service; DV as at least one 'most qualified' rating

VARIABLES	(1) Logit One or More 2LS MQ Extensive Margin Employee YOS = 3	(2) Logit One or More 2LS MQ Intensive Margin (Robust) Employee YOS = 3	(3) Logit One or More 2LS MQ Intensive Margin Narrow Employee YOS = 3	(4) Logit One or More 2LS MQ Extensive Margin (High Intensity) Employee YOS = 3	(5) Logit One or More 2LS MQ Int. Margin (Robust — High Intensity) Employee YOS = 3
IV: Extensive Margin Indicator (BH in 3 years) = 1	0.028			-0.172*	
	(0.031)			(0.099)	
IV: Robust Intensive Margin Indicator (BH in 3 years) = 1		0.020			
		(0.033)			
IV: Robust Intensive Margin Indicator (BH in 3 years) = 2		-0.019			-0.042
		(0.019)			(0.047)
IV: Narrow Intensive Margin Indicator (BH in 3 years) = 1			0.035		
			(0.051)		
1LS had more than mean BH appointments = 1					-0.118**
					(0.057)
Individual had BH appt in last 3 years = 1	-0.147***	-0.113***	-0.067***	-0.155***	-0.068***
	(0.029)	(0.021)	(0.019)	(0.030)	(0.019)
Indvidual had more than the mean BH appts (year) = 1	-0.203***	-0.164***	-0.087**	-0.215***	-0.088**
	(0.057)	(0.021)	(0.041)	(0.073)	(0.039)
All not white individuals (=1) = 1	-0.046***	-0.065***	-0.094***	-0.051***	-0.091***
	(0.015)	(0.009)	(0.023)	(0.016)	(0.024)
All females (=1) = 1	0.087***	0.086**	0.045	0.117***	0.043
	(0.033)	(0.041)	(0.061)	(0.023)	(0.058)
Individuals not married (=1) = 1	0.005	-0.017	-0.043**	0.001	-0.041**
	(0.011)	(0.013)	(0.018)	(0.013)	(0.018)
Rank Grouping = 2, 2	-0.062	-0.098	-0.293***	0.069	-0.299***
	(0.284)	(0.176)	(0.011)	(0.243)	(0.012)
Employee and supervisor are same gender (binary) = 1	-0.048	-0.014	0.032	-0.038	0.032
	(0.040)	(0.030)	(0.031)	(0.034)	(0.028)
Employee and supervisor are same race (binary) = 1	-0.005	0.014	0.052	-0.009	0.055
	(0.007)	(0.016)	(0.044)	(0.018)	(0.046)
Employee and supervisor match race and gender = 1	0.038*	0.014	-0.046**	0.049**	-0.049**
	(0.021)	(0.011)	(0.021)	(0.023)	(0.021)
Span of control (1LS), Logged	0.003	-0.014	-0.032	0.001	-0.035
	(0.012)	(0.019)	(0.030)	(0.009)	(0.031)
Span of control (2LS), Logged	0.045***	0.053***	0.060***	0.048***	0.060***
	(0.005)	(0.008)	(0.009)	(0.007)	(0.009)
Occupational workgroup = 2, Operations Support	-0.081***	-0.039***	-0.044***	-0.048***	-0.046***
	(0.012)	(0.009)	(0.011)	(0.011)	(0.011)
Occupational workgroup = 3, Force Sustainment	-0.040***	-0.019***	-0.037**	-0.020***	-0.036**
	(0.007)	(0.006)	(0.015)	(0.006)	(0.015)
Occupational workgroup = 4, Medical	0.027***	0.024***	-0.003	0.032***	-0.001
	(0.006)	(0.003)	(0.016)	(0.006)	(0.015)
Occupational workgroup = 5, Other	0.060***	0.048***	0.003	0.070***	0.003
	(0.006)	(0.001)	(0.013)	(0.006)	(0.013)
Total supervisor relationship length, Standardized	0.053***	0.047***	0.039	0.052***	0.040
	(0.016)	(0.013)	(0.024)	(0.016)	(0.025)
Presidential Administration = 2, Trump	0.007	0.012	0.011***	0.013	0.015***
	(0.020)	(0.013)	(0.004)	(0.023)	(0.003)
Presidential Administration = 3, Biden	-0.004	0.003	-0.009	0.005	-0.010
	(0.012)	(0.009)	(0.015)	(0.015)	(0.016)
Observations	2,848	4,456	1,859	2,625	1,859
Cluster	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with three years of service, and employees and supervisors who have sought BH in the last three years. The dependent variable is the probability that the employee has received one 'most qualified' rating on their annual evaluations at their rank. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes clustered standard errors based on employee occupational workgroup. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Table D.12: Regression results for employees with 7 years of service; DV as at least one 'most qualified' rating

VARIABLES	(1) Logit One or More 2LS MQ Extensive Margin Employee YOS = 7	(2) Logit One or More 2LS MQ Intensive Margin (Robust) Employee YOS = 7	(3) Logit One or More 2LS MQ Intensive Margin Narrow Employee YOS = 7	(4) Logit One or More 2LS MQ Extensive Margin (High Intensity) Employee YOS = 7	(5) Logit One or More 2LS MQ Int. Margin (Robust - High Intensity) Employee YOS = 7
IV: Extensive Margin Indicator (BH in 3 years) = 1	-0.072**				
	(0.030)				
IV: Robust Intensive Margin Indicator (BH in 3 years) = 1		-0.074**			
		(0.030)			
IV: Robust Intensive Margin Indicator (BH in 3 years) = 2		-0.015			0.073
		(0.029)			(0.047)
IV: Narrow Intensive Margin Indicator (BH in 3 years) = 1			-0.073		
			(0.047)		
1LS had more than mean BH appointments = 1					0.003
					(0.045)
Individual had BH appt in last 3 years = 1	-0.017	-0.033	-0.037**	-0.021	-0.037**
	(0.046)	(0.026)	(0.016)	(0.046)	(0.016)
Indvidual had more than the mean BH appts (year) = 1	-0.162**	-0.203***	-0.249***	-0.166***	-0.249***
	(0.076)	(0.060)	(0.064)	(0.060)	(0.069)
All not white individuals (=1) = 1	-0.033*	-0.029	-0.044	-0.028	-0.044
	(0.019)	(0.026)	(0.034)	(0.025)	(0.034)
All females (=1) = 1	0.026	0.032	0.049	0.030	0.049
	(0.060)	(0.054)	(0.075)	(0.055)	(0.075)
Individuals not married (=1) = 1	-0.070*	-0.097***	-0.176***	-0.054	-0.176***
	(0.039)	(0.019)	(0.031)	(0.042)	(0.031)
Rank Grouping = 2, 2	-0.283***	-0.292***	-0.272***	-0.303***	-0.272***
	(0.026)	(0.023)	(0.044)	(0.034)	(0.044)
Employee and supervisor are same gender (binary) = 1	-0.033	-0.046*	-0.030	-0.052	-0.030
	(0.029)	(0.024)	(0.071)	(0.037)	(0.072)
Employee and supervisor are same race (binary) = 1	-0.029	-0.058	-0.113*	-0.023	-0.113*
	(0.042)	(0.041)	(0.066)	(0.047)	(0.066)
Employee and supervisor match race and gender = 1	0.093**	0.089**	0.055	0.101**	0.055
	(0.040)	(0.037)	(0.088)	(0.043)	(0.088)
Span of control (1LS), Logged	0.018	0.022	0.014	0.023	0.014
	(0.018)	(0.029)	(0.054)	(0.025)	(0.055)
Span of control (2LS), Logged	0.013***	0.020*	0.033	0.011**	0.033
	(0.005)	(0.011)	(0.027)	(0.005)	(0.027)
Occupational workgroup = 2, Operations Support	0.231***	0.235***	0.296***	0.201***	0.296***
	(0.019)	(0.031)	(0.052)	(0.019)	(0.052)
Occupational workgroup = 3, Force Sustainment	0.134***	0.082***	0.025	0.121***	0.025
	(0.004)	(0.019)	(0.054)	(0.002)	(0.052)
Occupational workgroup = 4, Medical	0.113***	0.158***	0.238***	0.099***	0.238***
	(0.010)	(0.012)	(0.054)	(0.008)	(0.054)
Occupational workgroup = 5, Other	0.127***	0.103***	0.104**	0.102***	0.104**
	(0.003)	(0.017)	(0.051)	(0.003)	(0.050)
Total supervisor relationship length, Standardized	0.078**	0.074***	0.083***	0.069**	0.083***
	(0.032)	(0.012)	(0.023)	(0.033)	(0.024)
Presidential Administration = 2, Trump	0.094*	0.085*	0.080**	0.070	0.079*
	(0.049)	(0.047)	(0.038)	(0.049)	(0.041)
Presidential Administration = 3, Biden	0.021	0.020	0.062	-0.017	0.062
	(0.047)	(0.057)	(0.045)	(0.056)	(0.045)
Observations	1,031	1,519	581	938	581
Cluster	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with seven years of service, and employees and supervisors who have sought BH in the last three years. The dependent variable is the probability that the employee has received one 'most qualified' rating on their annual evaluations at their rank. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes clustered standard errors based on employee occupational workgroup. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Table D.13: Regression results for employees with 11 years of service; DV as at least one 'most qualified' rating

VARIABLES	(1) Logit One or More 2LS MQ Extensive Margin Employee YOS = 11	(2) Logit One or More 2LS MQ Intensive Margin (Robust) Employee YOS = 11	(3) Logit One or More 2LS MQ Intensive Margin Narrow Employee YOS = 11	(4) Logit One or More 2LS MQ Extensive Margin (High Intensity) Employee YOS = 11	(5) Logit One or More 2LS MQ Int. Margin (Robust — High Intensity) Employee YOS = 11
IV: Robust Intensive Margin Indicator (BH in 3 years) = 1		0.142			
		(0.102)			
IV: Robust Intensive Margin Indicator (BH in 3 years) = 2		0.011			-0.042
		(0.055)			(0.122)
IV: Extensive Margin Indicator (BH in 3 years) = 1	0.148*				
	(0.082)				
IV: Narrow Intensive Margin Indicator (BH in 3 years) = 1			0.037		
			(0.117)		
IV: Extensive Margin Indicator (BH in 3 years) = 0, omitted				-	
Individual had BH appt in last 3 years = 1	-0.072***	-0.055*	0.065	-0.092***	0.066
	(0.018)	(0.029)	(0.057)	(0.026)	(0.058)
Indvidual had more than the mean BH appts (year) = 1	0.225***	0.143**	0.182	0.251**	0.140
	(0.047)	(0.066)	(0.180)	(0.111)	(0.204)
All not white individuals (=1) = 1	-0.098	-0.090	-0.075	-0.111	-0.077
	(0.086)	(0.065)	(0.081)	(0.075)	(0.082)
All females (=1) = 1	0.121	0.115**	0.111	0.079	0.119
	(0.119)	(0.053)	(0.127)	(0.146)	(0.121)
Individuals not married (=1) = 1	-0.043	-0.219*	-0.558***	-0.055	-0.563***
	(0.084)	(0.124)	(0.130)	(0.090)	(0.131)
Rank Grouping = 2, 2	-0.096	-0.099	-0.042	-0.111*	-0.037
	(0.063)	(0.081)	(0.090)	(0.059)	(0.090)
Employee and supervisor are same gender (binary) = 1	0.047	0.072	0.171	0.043	0.182
	(0.066)	(0.091)	(0.244)	(0.073)	(0.226)
Employee and supervisor are same race (binary) = 1	-0.037	0.058	0.152	0.047	0.161
	(0.203)	(0.190)	(0.344)	(0.195)	(0.330)
Employee and supervisor match race and gender = 1	0.020	-0.107	-0.250	-0.075	-0.258
	(0.123)	(0.184)	(0.411)	(0.102)	(0.399)
Span of control (1LS), Logged	0.068	0.011	0.031	0.030	0.030
	(0.069)	(0.027)	(0.089)	(0.071)	(0.091)
Span of control (2LS), Logged	-0.030	-0.013	-0.008	-0.008	-0.005
	(0.033)	(0.031)	(0.072)	(0.038)	(0.077)
Occupational workgroup = 2, Operations Support	-0.276***	-0.135***	-0.008	-0.268***	-0.004
	(0.035)	(0.048)	(0.043)	(0.052)	(0.047)
Occupational workgroup = 3, Force Sustainment	-0.359***	-0.255***	-0.071	-0.417***	-0.063
	(0.012)	(0.008)	(0.117)	(0.010)	(0.127)
Occupational workgroup = 4, Medical	-0.291***	-0.126**	0.009	-0.234**	0.016
	(0.111)	(0.063)	(0.069)	(0.109)	(0.067)
Occupational workgroup = 5, Other	-0.254***	-0.161***	0.023	-0.279***	0.027
	(0.021)	(0.021)	(0.044)	(0.028)	(0.050)
Total supervisor relationship length, Standardized	0.101**	0.151***	0.201***	0.078*	0.202***
	(0.049)	(0.038)	(0.040)	(0.044)	(0.039)
Presidential Administration = 2, Trump	0.085	0.076	0.090	0.093	0.097
	(0.055)	(0.046)	(0.108)	(0.067)	(0.113)
Presidential Administration = 3, Biden	0.076*	-0.015	-0.133	0.051	-0.127
	(0.044)	(0.056)	(0.097)	(0.048)	(0.098)
1LS had more than mean BH appointments = 1					0.120
					(0.153)
Observations	184	265	97	168	97
Cluster	Yes	Yes	Yes	Yes	Yes
Time	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy	BH Seek Ever-Never Dummy
Equation Modeled	Eq. 1	Eq. 2	Eq. 2	Eq. 1	Eq. 2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table presents robustness regression results for employees with 11 years of service, and employees and supervisors who have sought BH in the last three years. The dependent variable is the probability that the employee has received one 'most qualified' rating on their annual evaluations at their rank. The independent variables include various demographic and behavioral health indicators, as well as supervisor relationship characteristics. The model includes clustered standard errors based on employee occupational workgroup. BH = Behavioral Health; 1LS = First Line Supervisor; 2LS = Second Line Supervisor.

Bibliography

Acosta, J., Becker, A., Cerully, J., Fisher, M., Martin, L., Vardavas, R., Slaughter, M., and Schell, T. (2014). Mental Health Stigma in the Military. RAND Corporation.

Ahuja, G. and Katila, R. (2001). Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. Strategic Management Journal, 22(3):197–220.

Apfelbaum, E. P., Norton, M. I., and Sommers, S. R. (2012). Racial Color Blindness: Emergence, Practice, and Implications. Current Directions in Psychological Science, 21(3):205–209.

Apfelbaum, E. P., Pauker, K., Sommers, S. R., and Ambady, N. (2010). In Blind Pursuit of Racial Equality? Psychological Science, 21(11):1587–1592.

Asch, B. J., Miller, T., and Malchiodi, A. (2012). A New Look at Gender and Minority Differences in Officer Career Progression in the Military. Technical report, RAND Corporation.

Åslund, O. and Skans, O. N. (2012). Do Anonymous Job Application Procedures Level the Playing Field? ILR Review, 65(1):82–107.

Azmat, G., Cuñat, V., and Henry, E. (2020). Gender Promotion Gaps: Career Aspirations and Workplace Discrimination.

Baker, M., French, E., and Ali, M. (2021). Insights into Ineffectiveness of Gender Equality and Diversity Initiatives in Project-Based Organizations. Journal of Management in Engineering, 37(3):04021013.

Bar-Haim, Y., Ziv, T., Lamy, D., and Hodes, R. M. (2006). Nature and Nurture in Own-Race Face Processing. Psychological Science, 17(2):159–163.

Barnes, C. M. and Wagner, D. T. (2023). Leading for human sustainability: An extension of Restricted Employee Sustainability Theory. Research in Organizational Behavior, 43:100197.

Barnes, C. M., Wagner, D. T., Schabram, K., and Boncoeur, D. (2023). Human Sustainability and Work: A Meta-Synthesis and New Theoretical Framework. Journal of Management, 49(6):1965–1996.

Behaghel, L., Crépon, B., and Le Barbanchon, T. (2015). Unintended Effects of Anonymous Résumés. American Economic Journal: Applied Economics, 7(3):1–27.

Benson, A. M., Li, D., and Shue, K. (2023). Potential and the Gender Promotion Gap. Academy of Management Proceedings, 2023(1):19580.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. The American Economic Review, 94(4):991–1013.

Bøg, M. and Kranendonk, E. (2011). Labor Market Discrimination of Minorities? Yes, but not in Job Offers.

Butler, D. and Denton, S. W. (2021). How Effective Are Blinding Concepts and Practices to Promote Equity in the Department of the Air Force? RAND Corporation.

Campbell, M., Auchterlonie, J. L., Andris, Z., Cooper, D. C., and Hoyt, T. (2023). Mental Health Stigma in Department of Defense Policies: Analysis, Recommendations, and Outcomes. Military Medicine, 188(5-6):e1171–e1177.

Cancio, A. S., Evams, T. D., and Maume, D. J. (1996). Reconsidering the declining significance of race: Racial differences in early career wages. American Sociological Review, 61(4):541.

Castilla, E. J. (2008). Gender, Race, and Meritocracy in Organizational Careers. American Journal of Sociology, 113(6):1479–1526.

Castilla, E. J. (2011). Bringing Managers Back In: Managerial Influences on Workplace Inequality. American Sociological Review, 76(5):667–694.

Castilla, E. J. and Benard, S. (2010). The Paradox of Meritocracy in Organizations. Administrative Science Quarterly, 55(4):543–576.

Castilla, E. J. and Ranganathan, A. (2020). The Production of Merit: How Managers Understand and Apply Merit in the Workplace. Organization Science, 31(4):909–935.

Connelly, B. L., Tihanyi, L., Crook, T. R., and Gangloff, K. A. (2014). Tournament Theory: Thirty Years of Contests and Competitions. Journal of Management, 40(1):16–47.

Corrigan, P. W. (2000). Mental health stigma as social attribution: Implications for research methods and attitude change. CLINICAL PSYCHOLOGY.

Cotter, D. A., Hermsen, J. M., and Vanneman, R. (2003). The Effects of Occupational Gender Segregation across Race. The Sociological Quarterly, 44(1):17–36.

Dalen, P. (2014). Talent Management in the Army: Rethinking the Block Check. https://thestrategybridge.org/the-bridge/2014/2/23/talent-management-in-the-army-rethinking-the-block-check.

Defense Health Agency (2018). DHA-PM 6010.13 MEPRS Volume 1.

Department of the Army (2020). Elimination of Department of the Army Photos, and Race, Ethnicity, and Gender Identification Data for Officer, Warrant Officer, and Enlisted Selection Boards (Updated).

DiBenigno, J. (2018). Anchored Personalization in Managing Goal Conflict between Professional Groups: The Case of U.S. Army Mental Health Care. Administrative Science Quarterly, 63(3):526–569.

DiBenigno, J. (2020). Rapid Relationality: How Peripheral Experts Build a Foundation for Influence with Line Managers. Administrative Science Quarterly, 65(1):20–60.

DiBenigno, J. and Kerrissey, M. (2020). Structuring mental health support for frontline caregivers during COVID-19: Lessons from organisational scholarship on unit-aligned support. BMJ Leader, 4(3):124–127.

Dimoff, J. K. and Kelloway, E. K. (2019). With a little help from my boss: The impact of workplace mental health training on leader behaviors and employee resource utilization. Journal of Occupational Health Psychology, 24(1):4–19.

DoD Office of the Actuary (2023). Statistical Report on the Military Retirement System.

Eagly, A. H. and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. Psychological Review, 109(3):573–598.

Edmonson, A. (1999). Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2):350–353.

Eliezer, D., Roberts, C., Adams, J., Esposito, J., Mantua, J., Mitchell, T., Oswald, C., Reutter, H., Wang, E., and Williams, A. (2024). Assessing the Impact of Removing Demographic Indicators from Military Selection Boards. IDA Paper P-33193, Institute for Defense Analysis, Alexandria, VA.

Esper, M. T. (2020). Press Release to Address DEI and EO in the Military Services. https://www.defense.gov/News/Releases/Release/Article/2275106/.

Evans, C. R., Leckie, G., Subramanian, S., Bell, A., and Merlo, J. (2024). A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). SSM - Population Health, 26:101664.

Fleming, W. J. (2024). Employee well-being outcomes from individual-level mental health interventions: Cross-sectional evidence from the United Kingdom. Industrial Relations Journal, 55(2):162–182.

Forbes, J. B. (1987). Early Intraorganizational Mobility: Patterns and Influences. Academy of Management Journal, 30(1):110–125.

Gaddis, S. M. (2015). Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market. Social Forces, 93(4):1451–1479.

Geibel, H. V., Rigotti, T., and Otto, K. (2022). It all comes back to health: A three-wave cross-lagged study of leaders' well-being, team performance, and transformational leadership. Journal of Applied Social Psychology, 52(7):532–546.

Goldin, C. and Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. American Economic Review, 90(4):715–741.

González Ramos, A. M., Navarrete Cortés, J., and Cabrera Moreno, E. (2015). Dancers in the Dark: Scientific Careers According to a Gender-Blind Model of Promotion. Interdisciplinary Science Reviews, 40(2):182–203.

Graddy-Reed, A., Lanahan, L., and Eyer, J. (2019). Gender discrepancies in publication productivity of high-performing life science graduate students. Research Policy, 48(9):103838.

Greenman, E. and Xie, Y. (2008). Double Jeopardy? The Interaction of Gender and Race on Earnings in the United States. Social Forces, 86(3):1217–1244.

Hammer, L. B., Dimoff, J., Mohr, C. D., and Allen, S. J. (2024). A Framework for Protecting and Promoting Employee Mental Health through Supervisor Supportive Behaviors. Occupational health science, 8(2):243–268.

Harmon-Jones, E. and Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In Harmon-Jones, E., editor, Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology (2nd Ed.)., pages 3–24. American Psychological Association, Washington.

Hesli, V. L., Lee, J. M., and Mitchell, S. M. (2012). Predicting Rank Attainment in Political Science: What Else Besides Publications Affects Promotion? PS: Political Science and Politics, 45(3):475–492.

Hu, S., Gentry, R. J., Quigley, T. J., and Boivie, S. (2023). Who's in the Driver's Seat? Exploring Firm-Level vs. CEO-Level Effects on Problemistic Search. Journal of Management, 49(4):1187–1212.

Hultqvist, J., Hensing, G., Björk, L., and Bertilsson, M. (2024). Managers' attitudes to depression and the association with their rating of how work capacity is affected in employees with common mental disorders. BMC Research Notes, 17:144.

Ito, T. A. and Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. Journal of Personality and Social Psychology, 85(4):616–626.

Judge, T. A. and Martocchio, J. J. (1995). THE ROLE OF FAIRNESS ORIENTATION AND SUPERVISOR ATTRIBUTIONS IN ABSENCE DISCIPLINARY DECISIONS. Journal of Business & Psychology, 10(1):115–137.

Kaaronen, R. O. (2018). A Theory of Predictive Dissonance: Predictive Processing Presents a New Take on Cognitive Dissonance. Frontiers in Psychology, 9.

Krause, A., Rinne, U., and Zimmermann, K. F. (2012). Anonymous job applications in Europe. IZA Journal of European Labor Studies, 1(1):5.

Kreiner, G., Mihelcic, C. A., and Mikolon, S. (2022). Stigmatized Work and Stigmatized Workers. Annual Review of Organizational Psychology and Organizational Behavior, 9:95–120.

Kristof-Brown, A. L., Zimmerman, R. D., and Johnson, E. C. (2005). CONSEQUENCES OF INDIVIDUALS' FIT AT WORK: A META-ANALYSIS OF PERSON–JOB, PERSON–ORGANIZATION, PERSON–GROUP, AND PERSON–SUPERVISOR FIT. Personnel Psychology, 58(2):281–342.

Landau, J. (1995). The Relationship of Race and Gender to Managers' Ratings of Promotion Potential. Journal of Organizational Behavior, 16(4):391–400.

Lazear, E. P. and Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. Journal of Political Economy, 89(5):841–864.

Lemoine, G. J., Hartnell, C. A., and Leroy, H. (2019). Taking Stock of Moral Approaches to Leadership: An Integrative Review of Ethical, Authentic, and Servant Leadership. Academy of Management Annals, 13(1):148–187.

List, J. A. (2001). Determinants of securing academic interviews after tenure denial: Evidence from a zero-inflated Poisson model. Applied Economics, 33(11):1423–1431.

Luthans, F. and Church, A. H. (2002). Positive Organizational Behavior: Developing and Managing Psychological Strengths [and Executive Commentary]. The Academy of Management Executive (1993-2005), 16(1):57–75.

Lytell, M. C., Hansen, M. L., and Calkins, A. (2023). Retention of Racial-Ethnic Minorities in the Regular Army. Number A2011-1 in Report. RAND, Arroyo Center, Santa Monica, CA.

Markham, W. T., Harlan, S. L., and Hackett, E. J. (1987). PROMOTION OPPORTUNITY IN ORGANIZATIONS:. Personnel and Human Resource Management, 5(6):223–287.

Maslach, C. (1982). Burnout, the Cost of Caring. Englewood Cliffs, N.J. : Prentice-Hall.

McGuffin, J. J., Riggs, S. A., Raiche, E. M., and Romero, D. H. (2021). Military and Veteran help-seeking behaviors: Role of mental health stigma and leadership. Military Psychology, 33(5):332–340.

McInnis, K. J. (2020). Congress, Civilian Control of the Military, and Nonpartisanship. Technical Report IF11566, Congressional Research Service.

Mendel, R., Kissling, W., Reichhart, T., Bühner, M., and Hamann, J. (2013). Managers' reactions towards employees' disclosure of psychiatric or somatic diagnoses. Epidemiology and Psychiatric Sciences, 24(2):146–149.

Military Leadership Diversity Commission (2010). MLDC Issue Paper 34 - Officer Promotion and Command Selection Process. Original source removed during goverment mandated retraction of DEI materials. Contact author for source document.

Minitrez, H. (2020). Elimination of DA Photo, Race, Ethnicity & Gender Info From Soldier Records. https://www.army.mil/article/236851/elimination.

Norton, M. I., Vandello, J. A., and Darley, J. M. (2004). Casuistry and Social Category Bias. Journal of Personality and Social Psychology, 87(6):817–831.

Parker, P. A. and Kulik, J. A. (1995). Burnout, self- and supervisor-rated job performance, and absenteeism among nurses. Journal of Behavioral Medicine, 18(6):581–599.

Peña, K., Hinsen, K., and Wilbur, M. (2017). Why Diversity Programs Fail – And How to Fix Them.

Prescott, R. E. (2020). Removing photographs may not give a bigger picture for selection and promotion boards. https://www.militarytimes.com/opinion/commentary/2020/08/29/removing-photographs-may-not-give-a-bigger-picture-for-selection-and-promotion-boards/.

Public Service Commission of Canada (2018). Anonymized Recruitment Pilot Project - Final report. https://www.canada.ca/en/public-service-commission/services/publications/Name-blind-recruitment-pilot-project.html.

Quillian, L., Pager, D., Hexel, O., and Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proceedings of the National Academy of Sciences, 114(41):10870–10875.

Richmond, M. K., Shepherd, J. L., Pampel, F. C., Wood, R. C., Reimann, B., and Fischer, L. (2014). Associations Between Substance Use, Depression, and Work Outcomes: An Evaluation Study of Screening and Brief Intervention in a Large Employee Assistance Program. Journal of Workplace Behavioral Health, 29(1):1–18.

Rivera, L. A. (2012). Hiring as Cultural Matching: The Case of Elite Professional Service Firms. American Sociological Review, 77(6):999–1022.

Rosenbaum, J. E. (1979). Tournament Mobility: Career Patterns in a Corporation. Administrative Science Quarterly, 24(2):220–241.

Schaufeli, W. B., Leiter, M. P., and Maslach, C. (2009). Burnout: 35 years of research and practice. Career Development International, 14(3):204–220.

Sheridan, J. E., Slocum, J. W., Buda, R., and Thompson, R. C. (1990). Effects of Corporate Sponsorship and Departmental Power on Career Tournaments. Academy of Management Journal, 33(3):578–602.

Snipp, C. M. and Cheung, S. Y. (2016). Changes in Racial and Gender Inequality since 1970. The Annals of the American Academy of Political and Social Science, 663:80–98.

Swift, S. A., Moore, D. A., Sharek, Z. S., and Gino, F. (2013). Inflated Applicants: Attribution Errors in Performance Evaluation by Professionals. PLoS ONE, 8(7):e69258.

Taliaferro, C. A. L. (2015). Understanding the Army Selection-Board Process.

Thomas, K. M. and Wise, P. G. (1999). Organizational Attractiveness and Individual Differences: Are Diverse Applicants Attracted by Different Factors? Journal of Business and Psychology, 13(3):375–390.

Tsui, A. S., Egan, T. D., O&amp, C. A., apos, and Reilly, I. I. I. (1992). Being different: Relational demography and organizational attachment. Administrative Science Quarterly, 37(4):549–580.

Tsui, A. S. and O'Reilly III, C. A. (1989). BEYOND SIMPLE DEMOGRAPHIC EFFECTS: THE IMPORTANCE OF RELATIONAL DEMOGRAPHY IN SUPERIOR-SUBORDINATE DYADS. Academy of Management Journal, 32(2):402–423.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, 185(4157):1124–1131.

US Army (2020). AR 600-8-29: Officer Promotions. Headquarters Department of the Army, Washington, D.C.

van Dijk, H., van Engen, M. L., and van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. Organizational Behavior and Human Decision Processes, 119(1):38–53.

van Vianen, A. E., Shen, C.-T., and Chuang, A. (2011). Person–organization and person–supervisor fits: Employee commitments in a Chinese context. Journal of Organizational Behavior, 32(6):906–926.

Wildman, S. M., Armstrong, M., Davis, A. D., and Grillo, T. (1996). Privilege Revealed: How Invisible Preference Undermines America. NYU Press.

Yang, Y., Obrenovic, B., Kamotho, D. W., Godinic, D., and Ostic, D. (2024). Enhancing Job Performance: The Critical Roles of Well-Being, Satisfaction, and Trust in Supervisor. Behavioral Sciences, 14(8):688.

Yavorsky, J. E. (2019). Uneven Patterns of Inequality: An Audit Analysis of Hiring-Related Practices by Gendered and Classed Contexts. Social Forces, 98(2):461–492.

Younkin, P. and Kuppuswamy, V. (2018). The Colorblind Crowd? Founder Race and Performance in Crowdfunding. Management Science, 64(7):3269–3287.